

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/200688673>

Perceptual and Physical Properties of Amplitude Fluctuation and their Musical Significance

Thesis · January 2001

CITATIONS

67

READS

526

1 author:



[Pantelis Vassilakis](#)

Loyola Marymount University

33 PUBLICATIONS 293 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Community Noise Mitigation for Outdoor Music Festivals [View project](#)



Curriculum design delivery & assessment [View project](#)

UNIVERSITY OF CALIFORNIA

Los Angeles

Perceptual and Physical Properties of Amplitude Fluctuation
and their Musical Significance

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Ethnomusicology

by

Panteleimon Nestor Vassilakis

2001

© Copyright by
Panteleimon Nestor Vassilakis
2001

The dissertation of Panteleimon Nestor Vassilakis is approved.

Peter M. Narins

A. Jihad Racy

Roger W. H. Savage

Gary A. Williams

Roger A. Kendall, Committee Chair

University of California, Los Angeles

2001

TABLE OF CONTENTS

<u>Chapter 1</u>	Introduction - Outline and Contributions of the Study.....	1
<u>1.1</u>	Introduction.....	1
<u>1.2</u>	Outline and Contributions of the Study.....	6
<u>Chapter 2</u>	Musical Relevance of Amplitude Fluctuation.....	8
<u>2.1</u>	Musical Expression and the Manipulation of Amplitude Fluctuation.....	8
<u>2.2</u>	The Middle Eastern Mijwiz.....	10
<u>2.3</u>	Indonesian Gongs.....	13
<u>2.4</u>	Bosnian Ganga Songs.....	17
<u>Chapter 3</u>	Amplitude Fluctuation versus Amplitude Modulation: Implementation Error, Adjustment, and Implications.....	21
<u>3.1</u>	Introduction - Definitions.....	21
<u>3.2</u>	The Difference between Degree of Amplitude Fluctuation and Amplitude Modulation Depth.....	26
<u>3.3</u>	Adjusting Amplitude Modulation Depth Implementation to Reflect Degree of Amplitude Fluctuation.....	32
<u>3.4</u>	Implications of the Proposed Adjustment to Perceptual Studies Manipulating Degree of Amplitude Fluctuation.....	37
<u>3.4.1</u>	Influence of Fluctuation Degree on the Roughness of AM Tones.....	38
<u>3.4.2</u>	Influence of Sound Pressure Level (<i>SPL</i>) on the Roughness of AM Tones.....	40
<u>3.4.3</u>	Roughness Threshold of AM Tones as a Function of Modulation Depth, Fluctuation Degree, and Modulation Frequency.....	41
<u>3.4.4</u>	Roughness of a Beating Tone-Pair.....	43
<u>3.4.5</u>	Discussion.....	44
<u>3.5</u>	Summary.....	45
<u>Chapter 4</u>	Amplitude Fluctuation and Sound: Background, Definitions, and Clarification of Terms.....	47
<u>4.1</u>	Wave Processes and Sound Waves in Air.....	47
<u>4.2</u>	Beating as a Sensation.....	52
<u>4.3</u>	Sound Waves and Beats.....	55

<u>Chapter 5</u>	Perceptual Manifestations of Amplitude Fluctuation: Problem Analysis, Suggestions.....	63
<u>5.1</u>	Beats and Amplitude Fluctuations.....	63
<u>5.2</u>	Sound Signals and Energy Content of Amplitude Fluctuations.....	66
<u>5.2.1</u>	Drawbacks of Two-Dimensional Signal Representations.....	66
<u>5.2.1.1</u>	Two-Dimensional Signals and Energy.....	66
<u>5.2.1.2</u>	Two-Dimensional Signals and Amplitude.....	70
<u>5.2.2</u>	Do Amplitude Fluctuations Carry Energy?.....	75
<u>5.3</u>	Roughness: Models Quantifying Roughness and their Drawbacks	83
<u>5.3.1</u>	Introduction.....	83
<u>5.3.2</u>	Models Quantifying Roughness and their Drawbacks.....	86
<u>5.4</u>	The Difference Tone.....	92
<u>5.4.1</u>	History, Physical Evidence, Explanations.....	92
<u>5.4.2</u>	The Pitch-Shift Effect.....	99
<u>5.5</u>	Dichotic Interference.....	105
<u>Chapter 6</u>	Theoretical Models - Hypotheses - Experiments.....	110
<u>6.1</u>	Introduction - Apparatus.....	110
<u>6.2</u>	Sound Wave Interference in Dichotic Experiments.....	113
<u>6.2.1</u>	Introduction.....	113
<u>6.2.2</u>	Experiment 1: Beating Unisons?.....	115
<u>6.2.3</u>	Experiments 2a, b, & c: Dichotic Interference/Cancellation?	123
<u>6.2.3.1</u>	Experiment 2a: Effect of Dichotic Phase Differences on Loudness.....	125
<u>6.2.3.2</u>	Experiment 2b: Effect of Dichotic Phase Differences on Direction.....	130
<u>6.2.3.3</u>	Experiment 2c: Effect of Dichotic Phase Differences on Stereo Image Width.....	138
<u>6.2.3.4</u>	Discussion and Implications of Experiments 2a, b, & c	144
<u>6.2.4</u>	Experiment 3: Dichotic Beats and Sound Localization Cues.	147
<u>6.2.5</u>	Summary.....	161
<u>6.3</u>	Energy Content of Amplitude Fluctuations – Alternative Signal Representation.....	164
<u>6.3.1</u>	Theoretical Value of the Amplitude Fluctuation Energy in Two-Component Complex Signals.....	164
<u>6.3.2</u>	Alternative Signal Representation Accounting for the Energy Content of Amplitude Fluctuations.....	170
<u>6.3.3</u>	Experiment 4: Testing the Theoretical Value of the Amplitude Fluctuation Energy in Two-Component Complex Signals.....	176
<u>6.3.4</u>	Summary.....	189

6.4	Roughness Estimation Model – Roughness Degrees and Harmonic Interval Ratings in the Western Musical Tradition.....	193
6.4.1	Roughness Estimation Model for Complex Spectra.....	193
6.4.2	Limitations of the Proposed Roughness Estimation Model....	198
6.4.2.1	Effect of Phase on Degree of Amplitude Fluctuation....	198
6.4.2.2	Effect of Envelope Asymmetry on Roughness.....	202
6.4.2.3	Roughness of Continuous Spectra.....	204
6.4.2.4	Summary.....	205
6.4.3	Experiment 5: Roughness Degrees and Harmonic Interval Dissonance Ratings.....	206
6.4.4	Summary.....	226
<u>Chapter 7</u>	Summary, Conclusions, and Suggestions for Further Study.....	229
7.1	Introduction.....	229
7.2	Degree of Amplitude Fluctuation versus Amplitude Modulation Depth.....	231
7.3	Theoretical and Experimental Examination of the Wave Energy Carried by Amplitude Fluctuations.....	233
7.4	Alternative Representation of Sound Waves for Better Illustration of Wave Energy Quantities.....	237
7.5	Sound Interference Products and Neural Wave Interaction	238
7.6	Auditory Roughness Estimation of Complex Spectra - Roughness Degrees and Harmonic Interval Dissonance Ratings within the Western Musical Tradition.....	241
7.7	Manipulation of Amplitude Fluctuation and Musical Expression.....	247
<u>Appendix</u>	Examples of Three-Dimensional Twisted Spiral Complex Signals.....	249
<u>Glossary</u>	261
<u>Index of Authors</u>	295
<u>References</u>	298

LIST OF TABLES

<u>Table 3.1</u>	Relationship between Degree of Amplitude Fluctuation and Amplitude Modulation Depth (1%-50%).....	34
<u>Table 3.2</u>	Relationship between Degree of Amplitude Fluctuation and Amplitude Modulation Depth (51%-100%).....	35
<u>Table 6.1</u>	List of Signals Used to Produce the Stimuli-Pairs for Experiment 1, Testing Possible Beating Due to Diplacusis.....	118
<u>Table 6.2</u>	Mean Responses (10 Subjects) for the 17 Stimuli-Pair Categories in Experiment 1, Collapsed across the Frequency Factor.....	119
<u>Table 6.3</u>	List of Stimuli Used to Produce the Stimuli-Pairs for Experiment 2a, Testing the Effect of Dichotic Phase Differences on Loudness.....	126
<u>Table 6.4</u>	Mean Responses (10 Subjects) for the 36 Stimuli-Pair Categories in Experiment 2a, Collapsed across the Frequency Factor.....	127
<u>Table 6.5</u>	List of Stimuli Used to Produce the Stimuli-Pairs for Experiment 2b, Testing the Effect of Dichotic Phase Differences on Direction.....	131
<u>Table 6.6</u>	Mean Responses (10 Subjects) for 36 Stimuli-Pairs in Experiment 2b (300Hz).....	132
<u>Table 6.7</u>	Mean Responses (10 Subjects) for 36 Stimuli-Pairs in Experiment 2b (1700Hz).....	135
<u>Table 6.8</u>	List of Stimuli Used to Produce the Stimuli-Pairs for Experiment 2c, Testing the Effect of Dichotic Phase Differences on Stereo Image Width.....	138
<u>Table 6.9</u>	Mean Responses (10 Subjects) for 64 Stimuli-Pairs in Experiment 2c...	139
<u>Table 6.10</u>	Mean Responses (10 Subjects) Comparing All Combinations between the Frequency Range and Phase Factors in Experiment 2c.....	141
<u>Table 6.11</u>	List of Stimuli Used in Experiment 3, Testing Possible Beats Due to Dichotic Frequency Differences.....	149

<u>Table 6.12</u>	Experiment and Model Stimuli for the Eight Categorization Tasks in Experiment 3.....	151
<u>Table 6.13</u>	Average Amplitude Values of the Interfering Sines and of the Amplitude Fluctuation Component in Experiment 4.....	186
<u>Table 6.14</u>	Frequency and Amplitude Values of the Components of the Complex Tones Used to Construct the Stimuli in Experiment 5.....	208
<u>Table 6.15</u>	Mean Roughness and Dissonance Ratings for 13 Harmonic Intervals versus roughness estimated by 3 roughness estimation models (2 Groups of 10 Subjects; Experiment 5).....	210
<u>Table 6.16</u>	Mean Roughness and Dissonance Ratings Grouped According to Statistical Significance by Tuckey HSD Post-Hoc Comparisons ($\alpha = 0.05$); Experiment 5.....	218
<u>Table 6.17</u>	Pearson r Correlations of Estimated Roughness to Observed Roughness and Observed Dissonance for each Roughness Estimation Model.....	224

LIST OF FIGURES

<u>Figure 1.1</u>	Illustration of a sound signal with (a) steady amplitude and (b) amplitude that fluctuates over time.....	2
<u>Figure 2.1</u>	Ali Jihad Racy performing on the <i>mijwiz</i>	11
<u>Figure 2.2</u>	A UCLA student practicing on a <i>bonang barung</i> , part of the <i>Kjai Mendung</i> Javanese <i>gamelan</i> 1958.....	16
<u>Figure 2.3</u>	UCLA's Balinese <i>gamelan</i> performing for a traditional dance.....	16
<u>Figure 2.4</u>	Members of UCLA's Balkan Music Ensemble in a <i>ganga</i> song.....	20
<u>Figure 3.1</u>	Signal and spectrum of an amplitude-modulated sine with $m = 1$, based on Eq. (3.1).....	24
<u>Figure 3.2</u>	Signal and spectrum of an amplitude-modulated sine with $m = 0.5$, based on Eq. (3.1).....	28
<u>Figure 3.3</u>	Error in the implementation of AF-degree that results from assuming AM-depth = AF-degree. Based on Eq. (3.6).....	30
<u>Figure 3.4</u>	Adjustment to the implementation of AF-degree, based on Eq.(3.8)	36
<u>Figure 3.5</u>	AM-depth (after Terhardt, 1974a) and AF-degree (adjusted using Eq. (3.6)) ratios that result in half the roughness.....	38
<u>Figure 3.6</u>	Roughness detectability threshold as a function of AM-depth, AF-degree, and modulation frequency.....	41
<u>Figure 4.1</u>	Three-dimensional signal of a mistuned octave between sines.....	62
<u>Figure 5.1</u>	Two one-second-long sine signals illustrating the problem of energy representation.....	68
<u>Figure 5.2</u>	Signal and spectrum of a modulated sine with AM-depth = 2 (200%) resulting in AF-degree = 1.3333 (133.33%).....	71
<u>Figure 5.3</u>	Signals and spectra resulting from gradual increase in AF-degree from 100% to 200% (AM-depth from 100% to ∞ %).....	73

<u>Figure 5.4</u>	Signals and spectra produced by a good and a misaligned gear (after Krishnappa, 1998: 706).....	81
<u>Figure 5.5</u>	Illustration of the <i>first pitch-shift effect</i>	100
<u>Figure 5.6</u>	Illustration of the <i>second pitch-shift effect</i>	101
<u>Figure 6.1</u>	Mean responses (10 subjects) comparing the degree of beating of monaural and binaural stimuli with the same AM-Depth (Experiment 1).....	120
<u>Figure 6.2</u>	Mean responses (10 subjects) comparing the degree of beating of the binaural stimulus with AM-depth = 0% to the degree of beating of monaural stimuli with various AM-depths. Experiment 1.....	122
<u>Figure 6.3</u>	Tree diagram resulting from cluster analysis of the means (10 subjects; Table 6.4) in Experiment 2a.....	128
<u>Figure 6.4</u>	Tree diagram resulting from cluster analysis of the means (10 subjects; Table 6.6, 300Hz) in Experiment2b.....	133
<u>Figure 6.5</u>	Tree diagram resulting from cluster analysis of the means (10 subjects; Table 6.7, 1700Hz) in Experiment2b.....	136
<u>Figure 6.6</u>	Tree diagram resulting from cluster analysis of the means (10 subjects; Table 6.10) in Experiment 2c.....	142
<u>Figure 6.7</u>	Computer interface for the categorization tasks in Experiment 3.....	150
<u>Figure 6.8</u>	Categorization results for the 300/1700Hz steady sines (ST) in Experiment 3.....	155
<u>Figure 6.9</u>	Categorization results for the 300/1700Hz left-to-right rotating sines (LR) in Experiment 3.....	156
<u>Figure 6.10</u>	Categorization results for the 300/1700Hz amplitude-modulated sines (AM) in Experiment 3.....	157
<u>Figure 6.11</u>	Categorization results for the two-component dichotic tones in Experiment 3.....	159
<u>Figure 6.12</u>	Graph of a vector $e^{i\omega t}$ with length A , rotating on the complex plane with angular frequency ω (Eq. (6.11)).....	171

<u>Figure 6.13</u>	Twisted-spiral representation of a two-component complex signal with 1Hz frequency difference between components and (a) unequal or (b) equal amplitudes.....	172
<u>Figure 6.14</u>	Alternative illustration of the twisted-signal hypothesis.....	173
<u>Figure 6.15</u>	Complex analytic signal derived from vibrational measurements on a machinery gear (after Krishinappa, 1998: 709).....	179
<u>Figure 6.16</u>	Block diagram of the Hilbert Transform Demodulation process extracting the envelope of a two-component complex signal and calculating its power spectrum (Experiment 4).....	181
<u>Figure 6.17</u>	Frequency Zoom analysis of the signal in Experiment 4.....	183
<u>Figure 6.18</u>	Hilbert Transform design for Experiment 4.....	184
<u>Figure 6.19</u>	Envelope signal of the complex signal in Experiment 4, extracted using Eq. (6.16).....	184
<u>Figure 6.20</u>	Frequency Zoom analysis of the envelope signal in Experiment 4...	185
<u>Figure 6.21</u>	Roughness curves plotting the observed roughness of a pair of sines (equal amplitudes) as a function of a) frequency separation and b) frequency of the lower tone (after Sethares, 1998: 45).....	194
<u>Figure 6.22</u>	Effect of phase on the degree of amplitude fluctuation.....	199
<u>Figure 6.23</u>	Effect of envelope symmetry on roughness.....	202
<u>Figure 6.24</u>	Estimated roughness for all 13 intervals in the chromatic scale starting at C4 (Experiment 5).....	211
<u>Figure 6.25</u>	Mean, standard deviation, and standard error values (10 subjects) of the observed roughness for all intervals in the chromatic scale starting at C4 (Experiment 5).....	215
<u>Figure 6.26</u>	Mean, standard deviation, and standard error values (10 subjects) of the observed dissonance for all intervals in the chromatic scale starting at C4 (Experiment 5).....	215

<u>Figure 6.27</u>	Mean observed roughness versus mean observed dissonance (10 subjects) for all intervals in the chromatic scale starting at C4 (Experiment 5).....	216
<u>Figure 6.28</u>	Estimated roughness (Helmholtz 1885) versus observed roughness and observed dissonance for all 13 intervals in the chromatic scale starting at C4 (Experiment 5).....	222
<u>Figure 6.29</u>	Estimated roughness (Hutchinson & Knopoff, 1978) versus observed roughness and observed dissonance for all 13 intervals in the chromatic scale starting at C4 (Experiment 5).....	223
<u>Figure 6.30</u>	Estimated roughness (proposed model, Eq. (6.23)) versus observed roughness and observed dissonance for all 13 intervals in the chromatic scale starting at C4 (Experiment 5).....	223
<u>Appendix</u>	Examples of three-dimensional spiral sine signals and twisted spiral complex signals..	249

LIST OF EQUATIONS

<u>Equation 3.1</u>	Sinusoidal amplitude modulation of a sine.....	23
<u>Equation 3.2</u>	Assumed relationship between minimum and maximum amplitudes of a modulated sine with modulation depth m	26
<u>Equation 3.3</u>	Degree of amplitude fluctuation as a function of the minimum and maximum amplitudes of a modulated signal with depth m	26
<u>Equation 3.4</u>	Actual relationship between minimum and maximum amplitudes of a modulated sine with depth m	27
<u>Equation 3.5</u>	Amplitude modulation depth as a function of the minimum and maximum amplitudes of a modulated signal with depth m	27
<u>Equation 3.6</u>	Degree of amplitude fluctuation as a function of amplitude modulation depth.....	29
<u>Equation 3.7</u>	Time-variant amplitude as a function of modulation frequency and modulation depth.....	32
<u>Equation 3.8</u>	Modulation depth coefficient values that need to be inserted in Equation 3.7 for an intended amplitude fluctuation degree to be applied.....	32
<u>Equation 3.9</u>	Roughness of an AM Sine as a function of degree of amplitude fluctuation.....	39
<u>Equation 3.10</u>	Roughness of an AM Sine as a function of sound pressure level...	40
<u>Equation 3.11</u>	Roughness of a sine-pair as a function of the roughness of an AM Sine.....	43
<u>Equation 5.1</u>	Intensity of a two-component complex wave as a function of the intensities of its two sine components.....	65
<u>Equation 5.2</u>	Sinusoidal amplitude modulation of a sine.....	70
<u>Equation 5.3</u>	Displacement as a function of time for two simple harmonic motions.....	75

<u>Equation 5.4</u>	A two-component signal with equal amplitudes considered as a sine (cosine) with slowly varying amplitude.....	75
<u>Equation 5.5</u>	Condition that needs to be met for a two-component signal to be considered as a sine (cosine) with slowly varying amplitude.....	75
<u>Equation 5.6</u>	Relationship between group and phase velocities of a modulated wave.....	76
<u>Equation 6.1</u>	A two-component signal with equal amplitudes considered as a sine with slowly varying amplitude.....	166
<u>Equation 6.2</u>	Equation 6.1 using angular frequencies.....	166
<u>Equation 6.3</u>	Power radiated by a point source pulsating at a single frequency....	167
<u>Equation 6.4</u>	Power radiated by a point source pulsating with the same amplitude at two different frequencies.....	167
<u>Equation 6.5</u>	Power radiated by a point source pulsating with the same amplitude at two different frequencies when considered as a single point source pulsating with time variant amplitude.....	168
<u>Equation 6.6</u>	Proposed model estimating the power radiated by amplitude fluctuations based on Equations 6.4 & 6.5.....	168
<u>Equation 6.7</u>	Power radiated by the amplitude fluctuations as a function of the power radiated by a point source pulsating with the same amplitude at two different frequencies.....	169
<u>Equation 6.8</u>	A two-component signal with unequal amplitudes considered as a sine with slowly varying amplitude.....	169
<u>Equation 6.9</u>	Time-dependent phase for Equation 6.8.....	169
<u>Equation 6.10</u>	Simple harmonic motion.....	170
<u>Equation 6.11</u>	Complex equation of a simple harmonic motion.....	171
<u>Equation 6.12</u>	Power radiated by the amplitude fluctuations, based on the wave parameters of the time-varying amplitude.....	174

<u>Equation 6.13</u>	Real signal (i.e. without imaginary terms) with frequency f and slowly varying amplitude $A(t)$	178
<u>Equation 6.14</u>	Complex (i.e. including imaginary terms) analytic (i.e. with a spectrum that has only positive frequency components) signal corresponding to the signal in Equation 6.13.....	178
<u>Equation 6.15</u>	Hilbert transform of the signal in Equation 6.13.....	179
<u>Equation 6.16</u>	Envelope of a signal as a function of the signal and its Hilbert transform.....	180
<u>Equation 6.17</u>	A two-component signal with equal amplitudes considered as a sine with slowly varying amplitude.....	182
<u>Equation 6.18</u>	Positively skewed Gaussian distribution.....	195
<u>Equation 6.19</u>	Roughness curves interpreted as positively skewed Gaussian distributions.....	195
<u>Equation 6.20</u>	Roughness as a function of amplitude fluctuation degree.....	196
<u>Equation 6.21</u>	Roughness as a function of sound pressure level.....	197
<u>Equation 6.22</u>	Roughness of a sine-pair as a function of the roughness of an amplitude modulated sine.....	197
<u>Equation 6.23</u>	Proposed model estimating the roughness of sine-pairs, based on Equations 6.19, 6.20, 6.21, & 6.22.....	197

ACKNOWLEDGMENTS

I would like to thank my family and friends for all their support and encouragement during this long process, especially my parents, Nestor and Koula Vassilakis, my brother Dimitri, my sister Kallirroï, and my editor and partner Maria.

A number of people consulted with and advised me during the course of this dissertation. I would like to thank especially John Hajda, Robert Reigle, Sonia Seeman, Angeles Sancho-Velazquez, Denis Claxton, Alice Hunt, and Kay Lubach.

I thank committee members Professors Ali Jihad Racy, Peter Narins, and Gary Williams for their comments and patience and Roger Savage for opening up my horizons to issues of philosophy and interpretation in music.

I owe thanks to the department of Ethnomusicology and to the Graduate Division at UCLA for supporting my pursuit of music education and recognizing my effort and work.

Finally, I would like to offer my gratitude and respect to Professor Roger Kendall for being a constant inspiration in my six years at UCLA, an exemplary mentor, and an overall unique person committed to teaching, research and above all music.

VITA

June 5, 1963	Born, Athens, Greece
1986-90	Composer, Producer Pulse Records Athens, Greece
1988-90	Photographer Athinaikes Publishing Athens, Greece
1992-93	Composer, Producer English National Ballet London, England
1993	B.A., Music Composition and Music Technology Kingston University Surrey, England
1993	London Arts Board Award for Original Cross-Cultural Contribution in the Arts London, England
1993-94	Composer, Producer London Chinese Orchestra London, England
1994	National Training Award London, England
1995	Sound Designer, Engineer BBC Radio 3 / London Greek Radio / Gateway Studios London, England
1997	M.A., Ethnomusicology University of California, Los Angeles

1996-99	Teaching Assistant, Research Assistant Department of Ethnomusicology University of California, Los Angeles
1999	First Prize, Gallery of Acoustics 138 th meeting of the Acoustical Society of America Columbus, Ohio
1999-2000	Annual Research Fellowship Graduate Division University of California, Los Angeles
2000	C.Phil., Ethnomusicology University of California, Los Angeles
2000	Instructor Department of Ethnomusicology University of California, Los Angeles
2000-01	Dissertation Year Fellowship Graduate Division University of California, Los Angeles

PUBLICATIONS AND PRESENTATIONS

- Vassilakis, P. (1997). First and second pitch-shift effects as alternative manifestations of a single phenomenon: introduction. *J. Acoust. Soc. Am.* 102 (5/2): 3186.
- Vassilakis, P. (1998). A single model explaining the first and second pitch-shift effects as alternative manifestations of a single phenomenon. *Proceedings of the 5th International Conference of Music perception and Cognition*: 107-114. Seoul, Korea: Seoul National University.
- Vassilakis, P. (1998). The first pitch-shift effect as a Doppler effect: a simple physical explanation to a complex perceptual phenomenon. *J. Acoust. Soc. Am.* 104 (3/2): 1799.

- Vassilakis, P. (1999). Chords as spectra, harmony as timbre. *J. Acoust. Soc. Am.* 106 (4/2): 2286 [supported by UCLA's Graduate division].
- Vassilakis, P. (1999). Those twisted beats. *Gallery of Acoustics*, First Prize. 138th meeting of the Acoustical Society of America. Columbus, Ohio.
- Vassilakis, P. (2000). Auditory roughness estimation of complex spectra. *Proceedings of the 6th International Conference of Music perception and Cognition*. CD-ROM. C. Woods, G. Luck, R. Brochard, F. Seddon, and J. A. Sloboda, editors. Keele University: Keele, England.
- Vassilakis, P. (2000). Amplitude modulation depth versus degree of amplitude fluctuation: implementation error, adjustment and implications. *J. Acoust. Soc. Am.* 108 (5/2): 2597.
- Vassilakis, P. (under development). *World Music*. CD-ROM presenting twelve music cultures with images, video clips, sound examples, and text providing historical, ethnographic, and organological information.

ABSTRACT OF THE DISSERTATION

Perceptual and Physical Properties of Amplitude Fluctuation and their Musical Significance

by

Panteleimon Nestor Vassilakis

Doctor of Philosophy in Ethnomusicology

University of California, Los Angeles, 2001

Professor Roger A. Kendall, Chair

Amplitude fluctuation (AF) is a manifestation of wave interference and a significant expressive tool in the production of musical sound. Musicians worldwide manipulate the AF parameters to exploit the sensations of beating and roughness, creating desired sonic effects. The history of the understanding of amplitude fluctuation is reviewed, revealing conflicting opinions regarding its physical properties and inconsistencies in its graphical representation. It is shown that degree of amplitude fluctuation (AF-degree) and amplitude modulation depth (AM-depth) are

conceptually and quantitatively different. Thus, previous studies manipulating AF-degree through implementation of AM-depth must be revised.

AFs satisfy all criteria defining a wave, including transfer of energy at the fluctuation rate. Theoretical values for the frequency and amplitude of the fluctuation component are compared to values obtained through Hilbert demodulation and frequency analysis, confirming the energy content of AF. However, the results are quantitatively inconclusive, warranting further study.

An alternative signal representation is proposed based on the complex equation of vibration, resulting in spiral sine signals and twisted-spiral complex signals. Three-dimensional spiral signals solve numerous problems associated with traditional signals and represent the energy content of AF.

The fact that AFs carry energy is challenged by previous experiments where sound interference products appear to arise at a neural level. New dichotic experiments illustrate that the perception of interference does not arise unless waves interact physically. The observed dichotic wave interaction matches results from sound localization studies.

Musical instrument construction and performance practice indicates that sound variations involving the sensation of roughness are found in most musical traditions. A new roughness estimation model is proposed demonstrating better agreement between estimated and observed roughness than earlier models. A hypothesis linking dissonance and roughness ratings of harmonic intervals within the chromatic scale is tested. Clear presence/absence of roughness appears to dominate dissonance ratings.

In other cases, dissonance decisions ignore roughness and are culturally/historically mediated. The results suggest that the consonance hierarchy of harmonic intervals in the Western musical tradition corresponds to variations in roughness degrees.

The study crosses disciplinary boundaries and improves the understanding of AF by examining musical practices worldwide. Further study should include cross-cultural empirical investigations.

[A]coustics is a "crossroads" subject. ... [It] bears its richest fruits when regarded as a synthesis of other classical disciplines rather than when pursued in textual isolation. ... That the problem of interdisciplinary communication is a real one, however, was illustrated not long ago when a distinguished contemporary physicist remarked casually, "I thought all the important problems in sound had been solved, and that this subject had been reduced to a matter of engineering practice." (Hunt, 1978: xiii-xiv, 6.)

Chapter 1 Introduction - Outline and Contributions of the Study

1.1 Introduction

Musical sounds are represented by vibration signals whose characteristics change with time. The term *amplitude fluctuation* describes variations in the maximum value (i.e. amplitude) of sound signals relative to a reference point¹ (Figure 1.1). As it will be argued (Chapters 3 to 5), the related literature contains conflicting opinions regarding the physical properties of amplitude fluctuation, as well as

¹ Defining amplitude as a maximum value of a vibration measured relative to a reference point rather than to a point of rest or to equilibrium is intentional and is addressed in Section 5.2.2.

inconsistencies in its graphical representation and manipulation (i.e. signals). As a result, there are similar inconsistencies regarding the possible perceptual attributes of amplitude fluctuation, hindering the understanding of its musical significance. The musical significance of amplitude fluctuation extends beyond the tuning considerations, which have often motivated its examination, and will be addressed in Chapter 2.

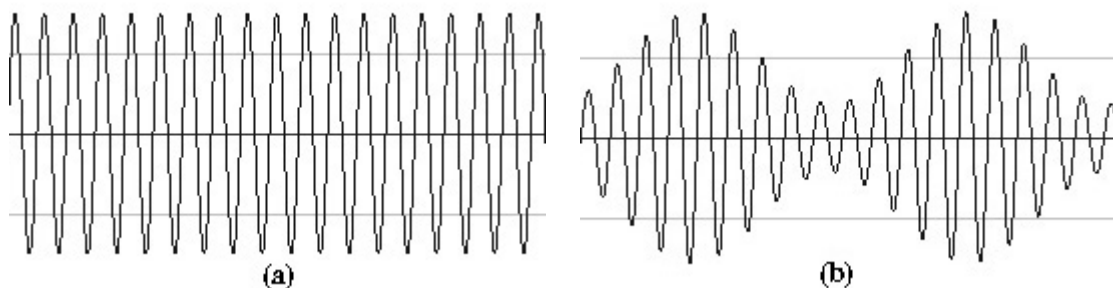


Figure 1.1: Illustration of a sound signal with (a) steady amplitude and (b) amplitude that fluctuates over time.

Amplitude fluctuations are manifestations of wave interference and can be placed in three overlapping perceptual categories related to the rate of fluctuation. Slow amplitude fluctuations ($\approx \leq 20$ per second) are perceived as loudness fluctuations referred to as *beating*. As the rate of fluctuation is increased, the loudness appears to be constant and the fluctuations are perceived as 'fluttering' or *roughness*.² As the

² The term roughness was first introduced in the acoustics and psychoacoustics literature by Helmholtz (1885) to refer to harsh, raspy, hoarse sounds.

amplitude fluctuation rate is increased further, the roughness reaches a maximum strength and then gradually diminishes until it disappears (≈ 75 -150 fluctuations per second, depending on the actual vibration frequency). These distinct perceptual categories do not reflect any fundamental qualitative differences in the vibrational frame of reference and should be approached as alternative manifestations of a single physical phenomenon.

Assuming the ear performs a frequency analysis on incoming signals,³ the above perceptual categories can be related directly to the bandwidth of the hypothetical analysis-filters (Zwicker, 1961). For example, in the simplest case of amplitude fluctuations resulting from the addition of two sine signals (f_1, f_2) ⁴, the fluctuation rate is equal to the frequency difference between the two sines $|f_1 - f_2|$ and the following statements represent the general consensus:

- If the filter-bandwidth is much larger than the fluctuation rate, then a single tone is perceived either with fluctuating loudness (i.e. beating) or with roughness.

³ As indicated by Ohm's acoustical law (Ohm, 1843, in Helmholtz, 1885, Plomp, 1965, etc.; Plomp, 1964).

⁴ A *sine signal* is a signal with a sinusoidal shape. Such a signal represents the simplest type of vibration (called simple harmonic motion - similar to a simple pendular motion) over time. The wave originating from a simple harmonic motion and represented by a sine signal is called *sine wave*. When a sine wave with frequency and amplitude values within the auditory limits reaches the ear, it gives rise to the sensation of a pure tone (i.e. similar to the sound produced by a tuning fork or an electric sine-wave generator).

- If the filter-bandwidth is much smaller than the fluctuation rate, then a complex tone⁵ is perceived, to which one or more pitches can be assigned but which, in general, exhibits no beating or roughness.⁶

In the first case, the degree, rate, and shape of a complex signal's amplitude fluctuations are variables that are manipulated by musicians of various cultures to exploit the beating and roughness sensations, making amplitude fluctuation a specific and significant expressive tool in the production of musical sound. In the second case (when there is no pronounced beating or roughness), degree, rate, and shape of a complex signal's amplitude fluctuations are variables that continue to be important

⁵ The term *complex tone* refers to the sensation arising from sound waves represented by complex signals. Any signal that does not have a sinusoidal shape is called *complex*. Fourier (early 1800s) proved that complex signals can be analyzed mathematically into the sum of a set of sinusoidal signals. These are called the Fourier components or *partials* of a given complex signal and make up the complex signal's *spectrum*. Analysis of a complex signal into sinusoidal components is called Fourier analysis, while the reverse process (i.e. constructing a complex signal out of a set of sinusoids) is called Fourier synthesis. For a periodic signal, the lowest frequency component is called the fundamental. The partials of a periodic signal have frequencies that are integer multiples of the frequency of the fundamental (i.e. if the fundamental has frequency f , then the partials have frequencies $1f, 2f, 3f, 4f, \dots$ etc.). Such a complex signal is called harmonic; all other signals are called inharmonic.

⁶ The beating and roughness sensations associated with certain complex tones are essentially understood in terms of sine-component interaction within the same *critical band*. The term *critical band*, introduced by Fletcher in the 1940s, referred to the frequency bandwidth of the, then loosely defined, auditory filter. Since von Békésy's studies (1960) the term also refers literally to the specific area on the basilar membrane (located in the inner ear, inside the cochlea) that goes into vibration in resonance with an incoming sine wave. Its length is determined by the elastic properties of the membrane and has an average value of $\sim 1\text{mm}$. However, studies by von Békésy (1960: 577-590) and Plomp (1966) examining the beating and roughness of mistuned consonances (i.e. sines with frequency ratios slightly removed from a small integer ratio) indicate that these sensations arise even when the added sines are separated by frequencies much larger than the critical bandwidth. The experimental results of both studies challenge earlier explanations of this phenomenon that were based on the nonlinear creation of combination tones (Helmholtz, 1885: 197-211) or harmonics (Wegel & Lane, 1924, in Plomp, 1966: 463) inside the ear. Although their final interpretations differ, both studies link the beating and roughness sensations of mistuned consonances directly to the resulting two-component complex signal's amplitude fluctuations.

through their interaction with the signal's spectral components. This interaction is manifested perceptually in terms of pitch or timbre variations linked to the introduction of combination tones. The majority of models constructed to account for this interaction proceed under the assumption that signal amplitude fluctuations represent pure patterns and transmit energy-lacking temporal messages, which translate into spectral patterns with no frequency component matching the fluctuation rate. Although there is a significant amount of experimental observations that clearly questions this assumption,⁷ a specific category of observations based on dichotic experiments⁸ is claimed to support it. Consequently, the issue of the physical and perceptual nature of amplitude fluctuation has remained unresolved, justifying the need for further study.

⁷ See Section 5.4.1.

⁸ In *dichotic* experiments, a different stimulus is presented to each ear. In *binaural* experiments, the same stimulus is presented to both ears. In *monaural* experiments, stimuli are presented to a single ear.

1.2 Outline and Contributions of the Study

The goal of this study is to improve the understanding of the perceptual and physical properties of amplitude fluctuation and better appreciate their musical significance. The need for such a study and its relevance to music are addressed in detail in Chapters 2 to 5. The study's main contributions are:

- I. Qualitative and quantitative analysis of the difference between amplitude fluctuation and amplitude modulation (Chapter 3).
- II. Theoretical and experimental determination of the wave energy carried by amplitude fluctuations (Sections 5.2, 5.4, and 6.3).
- III. Development of an alternative representation of sound waves that
 - a) represents consistently the energy of sine waves,
 - b) accounts for the possibility of negative amplitude values for certain complex signals, and
 - c) represents the energy content of amplitude fluctuations(Sections 5.2 and 6.3).

- IV. Examination of the possibility of sound interference products arising from wave interaction at levels other than the physical (i.e. neural) (Sections 5.5 and 6.2).
- V. Development of a model quantifying the roughness of complex spectra and examination of the relationship between roughness and dissonance ratings of harmonic intervals within the Western musical tradition (Sections 5.3 and 6.4).
- VI. Examination of ways in which the manipulation of amplitude fluctuation has been explored by various musical traditions in the production of musical sound (Chapter 2).

Chapter 2 Musical Relevance of Amplitude Fluctuation

2.1 Musical Expression and the Manipulation of Amplitude Fluctuation

Amplitude fluctuation constitutes the most common time-variant characteristic of musical sounds. The large variety of perceptual possibilities introduced through the manipulation of amplitude fluctuation (i.e. beating, roughness, combination tones) has been exploited by numerous musical traditions, a practice that has not yet been documented or researched. Manipulating the degree and rate of amplitude fluctuation helps create a shimmering (i.e. Indonesian *gamelan* performances), buzzing (i.e. Indian *tambura* drone), or rattling (i.e. Bosnian *ganga* singing) sonic canvas that becomes the backdrop for further musical elaboration. It permits the creation of timbral (i.e. Middle Eastern *mijwiz* playing) or even rhythmic (i.e. *ganga* singing) variations through gradual or abrupt changes between fluctuation rates and degrees.⁹ Whether such variations are explicitly sought for (as in *ganga* singing, *mijwiz* playing,

⁹ Such changes exploit the differences not only between the beating and roughness sensations but also among various degrees of roughness. Depending on the rate of fluctuation, three 'shades' of roughness have been distinguished (von Békésy, 1960: 354). Approximately 50 fluctuations per second give roughness of an intermediate character, lying between that of lower rates ("R" character) and that of higher rates ("Z" character).

or even the use of modulation wheels/pedals in modern popular music) or are introduced more subtly and gradually (as may be the case in the typical chord progressions/modulations of Western music), they form an important part of a musical tradition's expressive vocabulary. In North Indian music, for example, the buzzing sound of the *tambura* strings is considered as the life of a musical piece.¹⁰ It is based on a dynamic drone of up to four tones that interact with each other and with the melodic line they support (Carterette et al., 1989: 87). This interaction between background (drone) and foreground (melodic line) results in roughness variations that outline the function of each note within a *rāg* (scale) and set up patterns of tension and release within a specific piece of music (Jairazbhoy, 1995: 65).

Important clues regarding the ways various musical cultures approach the perceptual attributes of amplitude fluctuation can be found through an examination of musical instrument construction and performance practice. Additionally, the different choices among musical traditions with regards to vertical sonorities (i.e. harmonic intervals, chords, etc.) can reveal the variety of attitudes towards the sonic possibilities associated with the manipulation of amplitude fluctuation. The following Sections provide examples of musical traditions from around the world that use amplitude fluctuation as an expressive tool.

¹⁰ The *tambura* is an unfretted, long-necked lute with four strings (tuned to Octaves and a Fifth, a Fourth, or a Seventh, depending on tuning set), used to provide the drone accompaniment for one or more melodic instruments. A cotton thread (*juari*: life giving thread) is inserted between the bridge and the strings, giving the modulation of the string-length (as the string wraps/unwraps around the bridge) a discontinuous character and resulting in the characteristic, for the instrument, buzzing sound.

2.2 The Middle Eastern Mijwiz

The *mijwiz* is an aerophone made out of two identical cylindrical cane pipes, each with a single reed, bound together and played simultaneously (Figure 2.1). Double reed-pipes are found all across the Mediterranean. Their considerable cross-cultural and historical consistency with regards to playing technique, musical style, and even musical symbolism has often been attributed to their peculiar acoustical properties (Racy, 1994: 38).

The *mijwiz* has a stepped shape. Each of the two pipes are made by joining together three cane tubes of increasing diameter: one that has the vibrating reed, one that acts as a junction, and one that carries the tone-holes. This construction gives the instrument acoustical characteristics that resemble those of conical rather than cylindrical tubes, regarding register (an Octave higher than a cylinder of the same dimensions) and spectral composition (denser, richer spectra) of the produced tones. Additionally, the discontinuities of *mijwiz's* stepped design cause a complex mode distribution that results in slightly inharmonic spectral components, as is the case with all compound horns (Fletcher & Rossing, 1998: 217). The stepped design helps the instrument speak better because it allows for the use of a smaller reed system, as Racy has noted (1994: 42), but also because it provides better impedance matching between

the inside and the outside of the instrument than a cylindrical tube (Nederveen, 1998: 59-60). Nonetheless, since the effective cone-angle is very small, as is the diameter of the bore and of the tone holes, the resulting impedance matching is very weak. As a consequence, a large amount of pressure is required before the necessary standing waves can build up inside the tube and transfer energy outside the instrument, a pressure that must stay relatively constant if it is to sustain the free-reed vibrations and a steady tone.

The constant high pressure results not only in a *nasal* tone, rich in upper components, but also in an extremely limited dynamic range. Combined with the fact that the two reeds are activated by air pressure, with no manipulation possible from the lips or



Figure 2.1: Ali Jihad Racy performing on the *mijwiz*. Circular breathing is a technique used often by accomplished performers to ensure uninterrupted sound with no dynamic variation.

tongue, gives an instrument that has developed a celebrated¹¹ expressive power without relying on any of the usual sonic expressive tools. For its expressive power,

¹¹ Loaded with festive and ecstatic connotations, the *mijwiz* is hailed for inspiring "strong passions and exerting compelling powers and energies" (Racy, 1994: 50). The buzzing, rough sound of another double-pipe, the ancient Greek *aulos*, was prominent in cult rituals (Marcuse, 1975: 56).

the *mijwiz*, like most double-pipes throughout history, relies mainly on a single expressive tool: the manipulation and exploration of amplitude fluctuation rates.¹² The slight detuning between the two, otherwise identical, cane pipes (achieved through slight displacement of the tone-holes) means that when played together, they produce tones that beat constantly and at slightly shifting rates,¹³ giving the instrument a rich tonal quality. The amplitude fluctuation rate is explored further by occasionally increasing the detuning of the near-Unisons through partial stopping or by temporarily abandoning the Unisons and using one pipe as a high drone while performing a lower melody on the other. With a functional range of approximately a fifth and no possibility of line crossing, this technique represents a manipulation of roughness degrees rather than a form of polyphony. In its construction and performance practice, the *mijwiz* is an example of an instrument that makes explicit use of the perceptual richness of amplitude fluctuation, through creative exploration of the beating and roughness sensations.

¹² According to Jenkins & Olsen (1976: 58), the *mijwiz* and other double-pipes owe their exciting sound to beating.

¹³ The shifting beating rates are owed to the slight inharmonicity of the upper components and may be behind the "chorus effect" noted by Racy (1994: 44).

2.3 Indonesian Gongs

Gongs of various sizes and shapes are the backbone of Indonesian orchestras (*gamelan*).¹⁴ The mass, stiffness and thickness of the metal gong-shells play an important role in the tuning of the instrument's vibrational modes. Common to gongs of all sizes (Figure 2.2) are: a) a curved rim that creates an impedance mismatch between face and rim, essentially fixing the edges for most vibrational modes¹⁵ and b) a protruding (or sometimes sunken) dome/boss at the center, whose load is responsible for the Octave relationship between the two axially symmetric modes that are excited first and are associated with the gong's pitch. Numerous additional modes develop 200ms to 400ms after striking and decay at varying rates. These mostly higher modes are derived by the initial lower modes through bifurcation that eventually leads to a semi-chaotic vibrational behavior (similar in some respects to that of cymbals). The spectral energy shifts resulting from the above-mentioned bifurcation give rise to time-variant characteristics including amplitude fluctuation.

¹⁴ Some researchers trace the origin of East Asian gongs to the ancient Greek *ēcheion*, a large boss-less gong used in theaters as a resonator/amplifier (i.e. Marcuse, 1975: 52).

¹⁵ One of the main differences between gongs and other metallophones, such as bells, is that in gongs most vibratory energy emanates from the center rather than the edges of the instrument (Marcuse, 1975: 46). Fixing the edges facilitates the maintenance and control of beating effects within a single instrument.

These characteristics are equally chaotic since the related vibrational modes (bending modes) are not harmonically related and propagate on the gong-shell with dispersion.

In some cases, however, amplitude fluctuation becomes, intentionally, very regular. The *ombak* (waves/shimmer/beating) of the *gong agent*, the *penyorong* (regular beating) of the *pengisep* and *pengumbang* gong-pairs, and the beating of the *kajar* gong kettle are three such examples (Schneider & Beurmann, 1993). For the *gong agent*, the slow fluctuation and the resulting beating is caused by two nearby modes (modes 2 and 3; Harshberger, 1992: 60) being brought to the desired frequency relationship by the instrument makers through appropriate reshaping of the gong's body and boss (Marcuse, 1975: 48). Javanese musicians distinguish among four specific beating rates, and acoustical analysis has revealed *ombak* rates that range from 2 to 8 fluctuations per second (Carterette et al., 1993; Carterette and Kendall, 1994). For the *pengisep* and *pengumbang* gong-pairs, beating is the result of slight detuning of the two axially symmetric modes (which are the most significant in terms of energy content) between the two instruments. The exactness of the observed beating rates indicates expert intention behind instrument tuning. In the case of *kajar*, the sunken (rather than protruding) dome results in slower fluctuations than those observed in the sound of other small gongs (i.e. *bonang*), allowing the production of 'smoother' sounds (Schneider & Beurmann, 1993).

What is remarkable, in all cases, is the deep understanding of the acoustical properties of gongs by the instrument makers and musicians as well as the intentional manipulation of amplitude fluctuation rates to create specific sonic effects within a variety of performance contexts. In the East Javanese town of Banyuwangi, *seblang* fertility rituals involve trance states linked to the constantly beating backdrop of two *saron*¹⁶ (Wolbers, 1993). The Balinese *wayang gambuh* shadow-play has a distinct sonic character created by the beating among small bronze and brass idiophones (i.e. gong kettles such as the *kajar*, small cymbals, etc.). Sections of the music with pronounced beating support *tandakan* (melodic and textual improvisations) by the *dalang* (puppeteer) and heighten the listeners/spectators' awareness of the mythical dimension of *wayang* (Seebass, 1993).

Beating effects are encouraged by the acoustical properties and performance practice of gongs, most notably by the intentional out-of-tunedness of *gamelan*. As Becker (1988: 389) notes, imagining a context in which an out-of-tune ensemble is preferable to an in-tune one may be challenging for a Western listener but not for listeners that often value timbre over pitch. In the Javanese and Balinese musical traditions, timbres resulting from out-of-tune sound combinations may have special symbolic significance in that they constitute better representations of certain natural sounds (rain, thunder, oceans, volcanic eruption, etc.) than in-tune sound combinations. Consequently, an ensemble that can both represent and invoke natural

¹⁶ Type of metallophone. Modern *saron* are placed over wooden trough resonators (Marcuse, 1964, 457) that amplify and regulate any beating effects.

energies increases its metaphoric power by increasing its iconicity (Becker, 1988:389).

The intentional incorporation of specific beating effects in a variety of musical contexts indicates that, for the Javanese and Balinese musical traditions, the beating sound of gongs has both sonic and cultural significance.

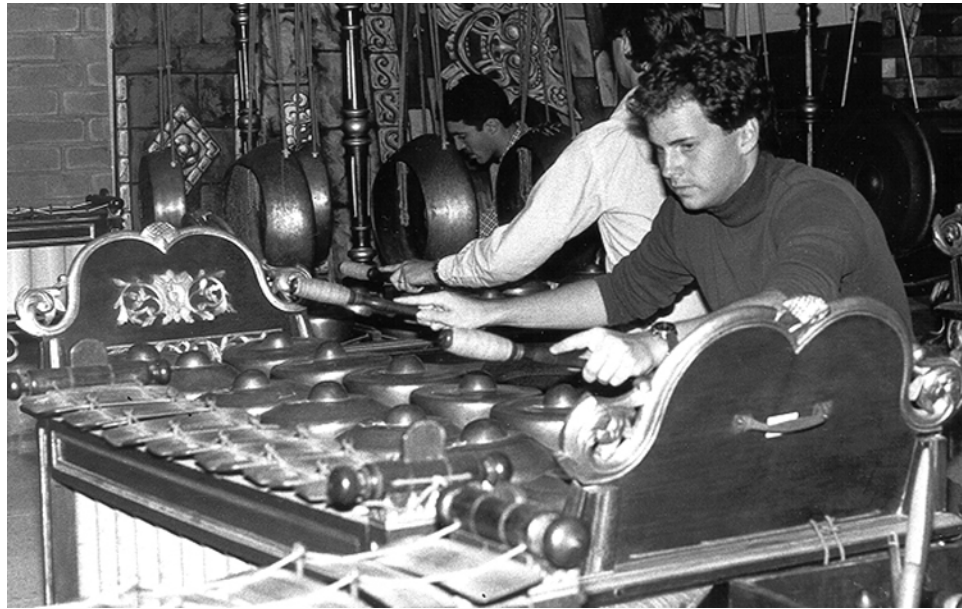


Figure 2.2: A UCLA student practicing on a *bonang barung* (tuned in *pelog*), part of the *Kjai Mendung* Javanese *gamelan* donated to UCLA by Mantle Hood in 1958.



Figure 2.3: UCLA's Balinese *gamelan* performing for a traditional dance.

2.4 Bosnian Ganga Songs

Ganga is a style of singing common in Bosnia-Herzegovina and the Dalmatian Zagora regions of the Balkans. *Ganga* songs consist of two alternating sections, one sung by a soloist and one by a soloist and a chorus (three to five singers). The melodic range rarely exceeds a fourth while, in the choral sections, voices sing at Minor / Major 2nd intervals that may or may not alternate with Unison passages. People in the region consider these intervals consonant and the resulting sound pleasant and desirable.¹⁷ Singing *ganga* provokes a feeling of corporate unity among singers and (initiated) listeners (Petrović, 1977: 336), and good performances have a strong emotional impact.

The length and content of the lyrics vary greatly among performances. The singers are relatively free in their choice of words, which are often nothing more than vocalizations (Petrović, 1977: 144), indicating that words are of far less importance to this genre than to most other types of folk songs. Within its geographical territory, *ganga* is valued for its distinct sonic effect rather than its semantic content. This

¹⁷ People in the specified regions generally associate the *ganga* sound with extreme joy. In contrast, the majority of outside listeners find the *ganga* sound annoying or even offensive.

effect relies mainly on the manipulation of and contrast between roughness degrees through (often rhythmic) alterations of solo, Unison, and Minor / Major 2nd passages.

The *ganga* style was initially approached by scholars as representing an inability to sing 'correctly' (Mariç, 1933, in Petroviç, 1977: 73). Further research revealed an explicit musical system, characterized by specific rules of musical creation and performance, and surprisingly fine (considering the narrow pitch and dynamic ranges employed) distinctions between sub-styles or good and bad songs / performances (Rihtman, 1951, in Petroviç, 1977: 76; Petroviç, 1977). Except for some stylized improvisation, *ganga* melodies are thoroughly composed and the principle structural feature of the songs is the contrast between solo and choral sections. This is simulated on a smaller scale in the choral passages through the alteration between Unisons and Minor / Major 2nds. The clear intention to achieve expressive goals through explicitly manipulating amplitude fluctuation is illustrated through the *ganga* rules governing a good song or performance:

- a) Singers must sing loudly and maintain uniform strength between each other and through time (Petroviç, 1977: 45, 101).
- b) The melodic range must not exceed a Major 3rd while harmonic intervals must not exceed a Major 2nd ¹⁸ (Petroviç, 1977: 326).

¹⁸ The few regions that have increased their melodic range to include wider harmonic intervals produce songs that are considered 'impure' and are referred to, with a negative connotation, as 'widely sung' (Petroviç, 1977: 260).

- c) The voices must be as identical as possible, nasal, and without vibrato so that they blend. Perfect blending of voices is a characteristic insisted upon heavily, requiring from the ensemble to ‘sound as one person’ (Petroviç, 1977: 308-309).

These loudness and timbral requirements are accompanied by very specific performance arrangements. Singers never move or dance while singing. They stand very tightly together, in an arch, turned slightly towards each other so that their voices will ‘collide’ at the right point (Figure 2.3). If those conditions are not fulfilled "the *ganga* will not be good." (Petroviç, 1977: 113). Performers refuse to sing under different conditions or to perform individual parts from the choral sections (when asked by researchers) since, when stripped from their perceptually rough intervals, these lines "make no sense" (Petroviç, 1977: 117). In other words, *ganga* represents a rare folk vocal genre where the sense of a song is related more to its sound than to its lyrics. Additionally, the conditions for a good song that makes sense are also conditions that guarantee the perceptual salience of amplitude fluctuations. This relationship is illustrated further by the three types of choral sections found in *ganga* songs:

- a) Two to three voice parts sing in Minor / Major 2nds with periodic insertions of Unisons, always resolving on a Major 2nd. The Unisons are not inserted randomly. Their function is to create specific rhythmic effects through the sudden contrast between perceptual roughness and smoothness. Petroviç

(1977: 288-294) includes a list of such rhythmic effects that are consistent across the *ganga* territory.

- b) Voices move in parallel seconds without Unison insertions.
- c) The leading (lower) part may cross the accompaniment and move a Major 2nd above it rather than remain, as is more often the case, below it.¹⁹

Although most songs start with a solo passage, performers refer to the entry of the choral section as the beginning of the song, and the importance of this section is closely related to its perceptually rough character.



Figure 2.4: Members of UCLA's Balkan Music Ensemble in a *ganga* song.

Ganga singing provides a striking example of a musical tradition with an expressive musical vocabulary that relies heavily on the sonic effects produced through the manipulation of amplitude fluctuation. Along with the examples sited previously, it indicates that a study on amplitude fluctuation is of interest to a musician as well as an engineer.

¹⁹ There are no cases of voices in contrary motion.

Chapter 3 Amplitude Fluctuation versus Amplitude Modulation: Implementation Error, Adjustment, and Implications

3.1 Introduction - Definitions

Chapter 3 identifies significant qualitative and quantitative differences between *amplitude fluctuation* and *amplitude modulation* and points to the problems associated with their confusion. The terms *amplitude fluctuation* and *amplitude modulation* are used interchangeably in the literature and the assumption of their equivalence is evident in definitions such as: "The difference in level between a signal's peak and the adjoining valley is referred to as modulation depth." (Levitt & Webster, 1998: 16.4). This and other such definitions make the specific assumption that degree of amplitude fluctuation (AF-degree) and amplitude modulation depth (AM-depth) are equivalent. The present chapter addresses the problems introduced by this erroneous assumption and introduces a possible solution.

The reason for the confusion between the two terms may be that the concept of modulation was introduced to signal processing from radio engineering. In radio engineering, the term *modulation* describes relatively slow variations of a signal's parameters (amplitude, frequency, and/or phase). More generally, modulations describe distortions of any arbitrary wave profile that become pronounced only on time intervals much longer than the period characteristic to the wave in question (called *carrier wave*). Modulations may be caused by the properties of the wave propagation medium or by the initial or boundary conditions. In any case, when seen from the reference point of the carrier, modulations can be considered waves rather than wave-distortions (Ostrovsky & Potapov, 1999: xiv).

As noted in the Federal Standard 1037C (1996), the term *amplitude modulation* also implies a specific spectral modification process²⁰ that produces discrete upper and lower sidebands (i.e. for sinusoidal carrier and modulating signals, the sum and difference frequencies of the carrier and the modulating signal). The envelope²¹ of the resulting modulated signal is then an analog of the modulating signal. Amplitude modulation (AM) of a sine (with frequency f , amplitude A , and phase ϕ) is therefore a process that introduces two additional components (sidebands)

²⁰ Mixing the carrier signal with the modulating signal in a nonlinear device.

²¹ The *envelope* of a signal is a curve that traces the signal's amplitude through time. It can be considered as the signal of the amplitude fluctuations of a modulated signal.

at frequencies determined by the modulation frequency f_{mod} and amplitudes determined by the modulation depth m . More specifically, modulation depth m indicates that the sum of the sidebands' amplitudes will be equal to m multiplied by the amplitude of the carrier. Eq. (3.1) gives the mathematical expression of the above definition:

$$(A + Am \cos 2\pi f_{\text{mod}} t) \sin(2\pi f t + \phi) = A \frac{m}{2} \sin[2\pi(f - f_{\text{mod}})t + \phi_1] + A \sin(2\pi f t + \phi) + A \frac{m}{2} \sin[2\pi(f + f_{\text{mod}})t + \phi_2] \quad \text{Eq. (3.1)}$$

where f_{mod} and m ($0 \leq m \leq \infty$) are the modulation parameters and f , A , and ϕ are the parameters of the sine with $|\phi_1 - \phi| = |\phi_2 - \phi| = \pi$ rads ($\pi = 3.14$; t : time).

Figure 3.1 shows the familiar spectrum and signal of an amplitude-modulated sine with depth $m = 1$. As it will be demonstrated, $m = 1$ is the only case where AF-degree and AM-depth are equivalent.

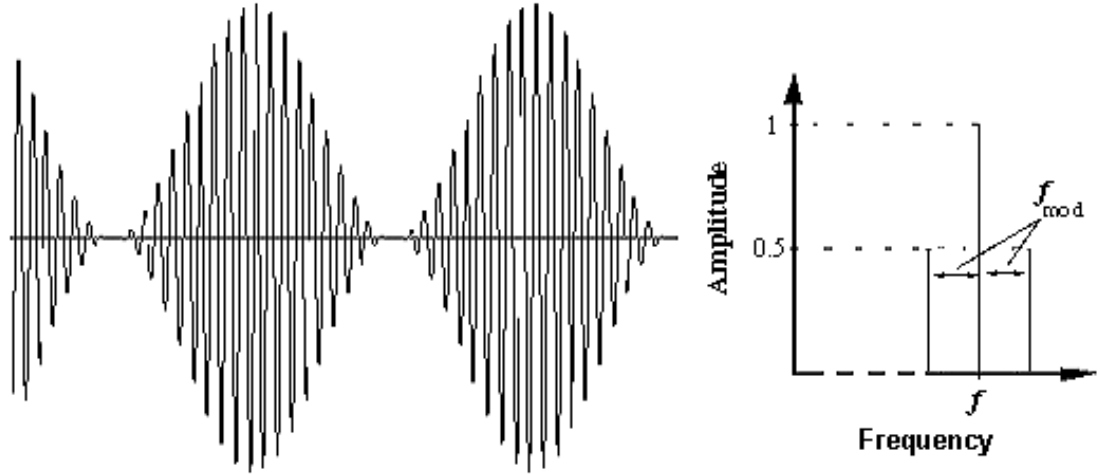


Figure 3.1: Signal and spectrum of an amplitude-modulated sine (carrier) with $m = 1$, based on Eq. (3.1). The spectrum includes the carrier frequency f and two sidebands $f - f_{\text{mod}}$ & $f + f_{\text{mod}}$, at amplitudes reflecting that modulation depth $m = 1$.

Figure 3.1 illustrates the intended spectral effect of the modulation process along with the side effect of amplitude fluctuations between a maximum (A_{max}) and a minimum (A_{min}) value at a rate equal to f_{mod} . The term side effect is chosen to emphasize that amplitude modulation depth, as implemented by Eq. (3.1), reflects (and is reflected on) spectral and not envelope considerations. Although it is a clear linear measure of spectral distribution, modulation depth is not a good measure of the level difference between a signal's peak and adjoining valley because it is qualitatively and quantitatively different from the degree of amplitude fluctuation.

The difference between AM-depth and AF-degree presents no problem as long as AM-depth is understood in terms of spectrum, the context it was defined in. However, tone/music perception studies use the term *amplitude modulation* to refer to the amplitude fluctuations of an AM tone's signal. Such studies are interested in the level difference between a signal peak and an adjoining valley (i.e. degree of amplitude fluctuation) and how it relates to a perceptual variable (i.e. pitch, residue pitch, timbre, roughness, or loudness of an AM tone, etc.) and not in modulation depth. Section 3.2 identifies the conceptual and quantitative differences between amplitude modulation depth and degree of amplitude fluctuation.

3.2 The Difference between Degree of Amplitude Fluctuation and Amplitude Modulation Depth

The underlying assumption in the vast majority of music perception studies that address amplitude modulation is that modulating the amplitude A of a sine with modulation frequency f_{mod} and depth m results in a signal whose amplitude fluctuates with rate f_{mod} between two values, A_{max} and A_{min} , so that:

$$A_{\text{max}} - A_{\text{min}} = mA_{\text{min}} \quad \text{or} \quad A_{\text{min}} = A_{\text{max}}(1 - m) \quad \text{Eq. (3.2)}$$

For example, applying a modulation depth of 0.4 (40%) is expected to result in a signal whose amplitude fluctuates between two values, A_{max} and A_{min} , so that: if

$A_{\text{max}} = 1$ then $A_{\text{min}} = 1(1 - 0.4) = 0.6$. Solving Eq. (3.2) for m gives:

$$m = \frac{A_{\text{max}} - A_{\text{min}}}{A_{\text{max}}} = AF_{\text{degree}} \quad \text{Eq. (3.3)}$$

Eq. (3.3) describes the **degree of amplitude fluctuation** AF_{degree} as a function of the minimum and maximum amplitudes of a modulated signal with depth m . If, however, we solve Eq. (3.1) accordingly, we get

$$A_{\max} = A + Am, \quad A_{\min} = A - Am, \quad \text{and therefore: } A_{\min} = A_{\max} \frac{1-m}{1+m} \quad \text{Eq. (3.4)}$$

Eq. (3.4) demonstrates that, in the case of an AM signal with depth m , the actual relationship among A_{\min} , A_{\max} , and m is not the one expected (Eq. (3.2)). Solving Eq. (3.4) for m gives:

$$m = \frac{A_{\max} - A_{\min}}{A_{\max} + A_{\min}} = AM_{depth} \quad \text{Eq. (3.5)}$$

Eq. (3.5) describes the **amplitude modulation depth** $AM_{depth} = m$ as a function of the minimum and maximum amplitudes of the modulated signal.

The difference between Eqs. (3.3) & (3.5) reflects a conceptual difference between degree of amplitude fluctuation (Eqs. (3.2) & (3.3)) and amplitude modulation depth (Eqs. (3.4) & (3.5)) and reveals a significant quantitative difference. For example, $AM_{depth} = 0.4$ (40%) indicates that the energy of the sidebands will be equal to 40% the energy of the carrier, but it does not indicate that the degree of

amplitude fluctuation will be equal to 40%. If $AM_{depth} = 0.4$ (40%) is inserted in Eq. (3.4) and we assume a sine with amplitude $A = 1$, then the resulting signal's amplitude will fluctuate between the values: $A_{max} = A + 0.4A = 1.4$ and $A_{min} = A - 0.4A = 0.6$. These values do not represent the intended 40% degree of amplitude fluctuation since A_{min} (0.6) is not equal to 60% of A_{max} (1.4) but equal to 43% of A_{max} (1.4). The degree of amplitude fluctuation is therefore $AF_{degree} = 100\% - 43\% = 57\% = 0.57$ and not 0.40. In other words, modulation depth (as implemented by Eq. (3.1)) is not a good measure of a signal's degree of amplitude fluctuation. For all values between 0% and 100%, AM-depth underestimates AF-degree. Figure 3.2 illustrates this point for the case of $AM_{depth} = 0.5$ (50%).

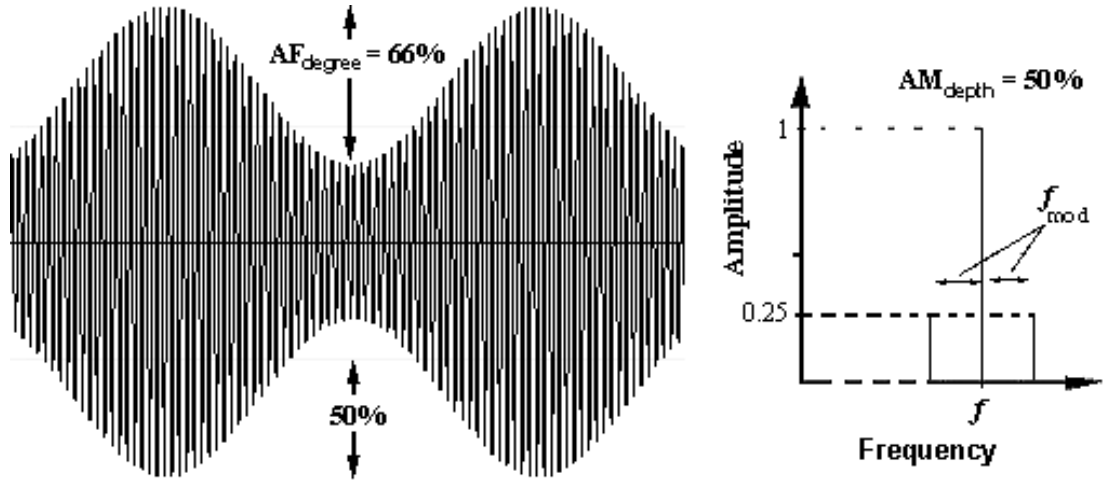


Figure 3.2: Signal and spectrum of an amplitude-modulated sine (carrier) with $m = 0.5$, based on Eq. (3.1). $AM_{depth} = 0.5$ (50%) results in a 3-component spectrum in which the sum of the sidebands' amplitudes is half the amplitude of the carrier. It also results in a signal whose amplitude fluctuates with $AF_{degree} = 0.66$ (66%) and not $AF_{degree} = 0.5$ (50%).

Combining Eqs. (3.3) & (3.5) gives an equation that links degree of amplitude fluctuation AF_{degree} to amplitude modulation depth AM_{depth} and quantifies the error resulting from assuming $AF_{degree} = AM_{depth}$:

$$AF_{degree} = \frac{2 AM_{depth}}{1 + AM_{depth}} \quad \text{Eq. (3.6)}$$

Based on Eq. (3.6), if $AM_{depth} = 0.4$ then $AF_{degree} = 0.57$. In other words, setting $m = 0.4$ in Eq. (3.1) will result in $AM_{depth} = 0.4$ but $AF_{degree} = 0.57$. Eq. (3.6) also indicates that:

- a) $AF_{degree} = AM_{depth}$ only if $AM_{depth} = 0$ or 1 : The assumption of equivalence between AM-depth and AM-degree is correct only for 0% & 100% modulation depths.
- b) If $AM_{depth} \rightarrow \infty$ then $AF_{degree} \rightarrow 2$: The upper limit of AF-degree is 2 (200%) instead of ∞ .

Figure 3.3 is a graphic representation of the error described by Eq. (3.6).

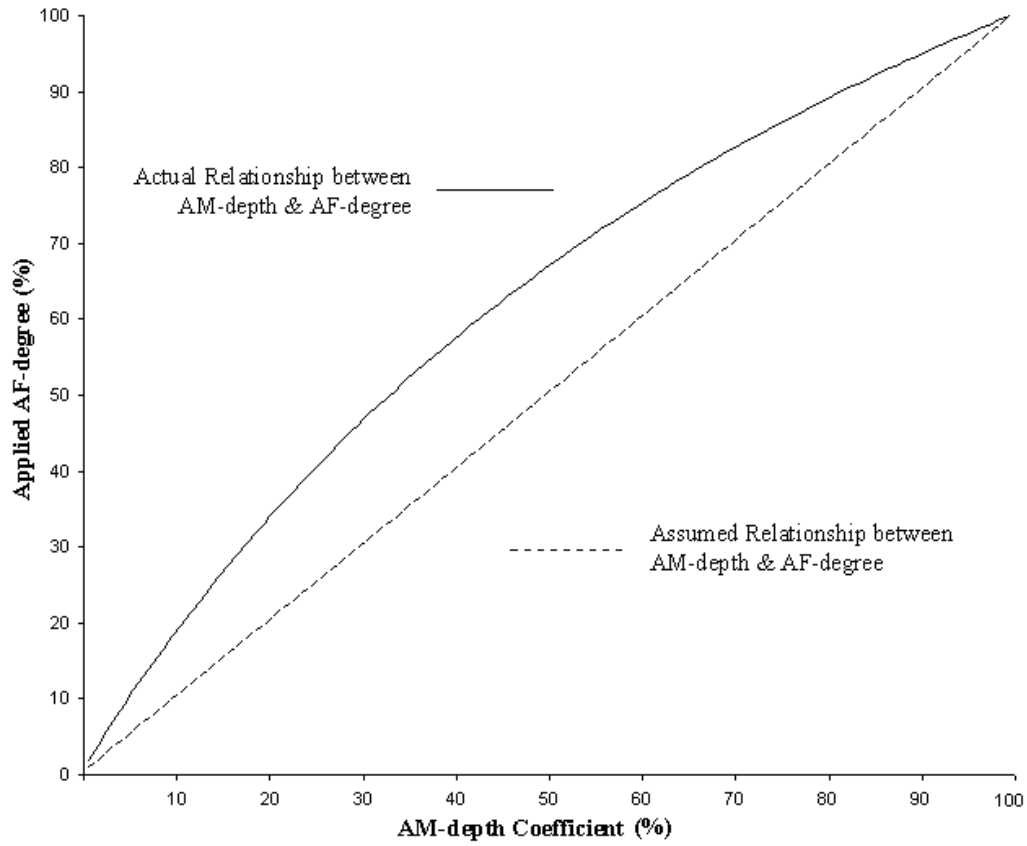


Figure 3.3: Error in the implementation of AF-degree that results from assuming AM-depth = AF-degree. Based on Eq. (3.6).

If our intention is to track envelope rather than spectral changes, manipulating a signal through manipulation of its AM-depth (Eq. (3.5)) (i.e. Iwamiya & Fujiwara, 1985; Iwamiya & Okamoro, 1996) will result in the discussed error. Eq. (3.5) is not an envelope-based definition of AM-depth so the changes it effects on the resulting signal's envelope are not systematic. Since, in the majority, perceptual studies are concerned with the perceptual nature of amplitude fluctuation of AM tones, Eq. (3.3)

is more useful than Eq. (3.5). It offers a precise, linear measure of the degree of amplitude fluctuation and, more importantly, it describes the precise effect that AM-depth is expected to have on the envelope of an amplitude-modulated signal.

The following Section corrects the error (Eq. (3.6)) that occurs when AM-depth is implemented through Eq. (3.1) to manipulate a signal's AF-degree. An adjusted coefficient m is calculated, which needs to be inserted in Eq. (3.1) in order for this equation to apply a specific degree of amplitude fluctuation rather than a specific degree of spectral energy spread. The implications of this adjustment to perceptual studies using AM tones are examined in Section 3.4.

3.3 Adjusting Amplitude Modulation Depth Implementation to Reflect Degree of Amplitude Fluctuation

Until now, regardless of whether AM-depth has been used to manipulate spectral energy spread or degree of amplitude fluctuation, amplitude modulation has always been implemented in the same manner:

$$A(t) = A + Am(\cos 2\pi f_{\text{mod}} t) \quad \text{Eq. (3.7)}$$

where f_{mod} and m ($0 \leq m \leq \infty$) are the modulation parameters; A : amplitude of the sine; t : time; $A(t)$: time variant amplitude ($\pi = 3.14$; zero initial phase).

Based on Eq. (3.6) we can derive a formula that calculates the modulation-depth coefficient m that need to be inserted in Eq. (3.7) for an intended degree of amplitude fluctuation AF_{degree} to be applied:

$$m = \frac{AF_{\text{degree}}}{2 - AF_{\text{degree}}} \quad \text{Eq. (3.8)}$$

For example, if the intended degree of amplitude fluctuation is $AF_{degree} = 0.4$ (40%), then the AM-depth coefficient we need to insert in Eq. (3.7) is not $m = 0.4$ but

$$m = \frac{0.4}{2 - 0.4} = 0.25 \text{ (25\%).}$$

Tables 3.1 & 3.2 demonstrate with numerical examples the relationship among:

- a) Intended degree of amplitude fluctuation AF_{degree} .
- b) Miscalculated/applied degree of amplitude fluctuation $AF_{degreeerror}$ if $m = AF_{degree}$ is inserted in Eq. (3.7).
- c) Adjusted AM-depth coefficient m that needs to be inserted in Eq. (3.7) for the intended degree of amplitude fluctuation to be applied.

All values are given in percentages, as is customary for AM-depth.

<p>Table 3.1 Relationship between Degree of Amplitude Fluctuation and Amplitude Modulation Depth (1%-50%) <i>AF_{degree}%</i>: Intended degree of amplitude fluctuation in 1% increments. <i>AF_{degree} error%</i>: applied degree of amplitude fluctuation assuming <i>m</i> = <i>AF_{degree}</i>; based on Eq. (3.6). <i>m %</i>: AM-depth coefficient that must be inserted in Eq. (3.7) for the intended degree of amplitude fluctuation <i>AF_{degree}</i> to be applied; based on Eq. (3.8).</p>					
<i>AF_{degree}</i>	<i>AF_{degree} error</i>	<i>m</i>	<i>AF_{degree}</i>	<i>AF_{degree} error</i>	<i>m</i>
1	1.980198	0.502513	26	41.26984	14.94253
2	3.921569	1.010101	27	42.51969	15.60694
3	5.825243	1.522843	28	43.75	16.27907
4	7.692308	2.040816	29	44.96124	16.95906
5	9.52381	2.564103	30	46.15385	17.64706
6	11.32075	3.092784	31	47.32824	18.3432
7	13.08411	3.626943	32	48.48485	19.04762
8	14.81481	4.166667	33	49.62406	19.76048
9	16.51376	4.712042	34	50.74627	20.48193
10	18.18182	5.263158	35	51.85185	21.21212
11	19.81982	5.820106	36	52.94118	21.95122
12	21.42857	6.382979	37	54.0146	22.69939
13	23.00885	6.951872	38	55.07246	23.45679
14	24.5614	7.526882	39	56.11511	24.2236
15	26.08696	8.108108	40	57.14286	25
16	27.58621	8.695652	41	58.15603	25.78616
17	29.05983	9.289617	42	59.15493	26.58228
18	30.50847	9.89011	43	60.13986	27.38854
19	31.93277	10.49724	44	61.11111	28.20513
20	33.33333	11.11111	45	62.06897	29.03226
21	34.71074	11.73184	46	63.0137	29.87013
22	36.06557	12.35955	47	63.94558	30.71895
23	37.39837	12.99435	48	64.86486	31.57895
24	38.70968	13.63636	49	65.77181	32.45033
25	40	14.28571	50	66.66667	33.33333

Table 3.2
Relationship between Degree of Amplitude Fluctuation and Amplitude Modulation
Depth (51%-100%)

AF_{degree}%: Intended degree of amplitude fluctuation in 1% increments.

AF_{degree error}%: applied degree of amplitude fluctuation assuming $m = AF_{degree}$; based on Eq. (3.6).

m %: AM-depth coefficient that must be inserted in Eq. (3.7) for the intended degree of amplitude fluctuation *AF_{degree}* to be applied; based on Eq. (3.8).

<i>AF_{degree}</i>	<i>AF_{degree error}</i>	<i>m</i>	<i>AF_{degree}</i>	<i>AF_{degree error}</i>	<i>m</i>
51	67.54967	34.22819	76	86.36364	61.29032
52	68.42105	35.13514	77	87.00565	62.60163
53	69.28105	36.05442	78	87.64045	63.93443
54	70.12987	36.9863	79	88.26816	65.28926
55	70.96774	37.93103	80	88.88889	66.66667
56	71.79487	38.88889	81	89.50276	68.06723
57	72.61146	39.86014	82	90.10989	69.49153
58	73.41772	40.84507	83	90.71038	70.94017
59	74.21384	41.84397	84	91.30435	72.41379
60	75	42.85714	85	91.89189	73.91304
61	75.7764	43.88489	86	92.47312	75.4386
62	76.54321	44.92754	87	93.04813	76.99115
63	77.30061	45.9854	88	93.61702	78.57143
64	78.04878	47.05882	89	94.17989	80.18018
65	78.78788	48.14815	90	94.73684	81.81818
66	79.51807	49.25373	91	95.28796	83.48624
67	80.23952	50.37594	92	95.83333	85.18519
68	80.95238	51.51515	93	96.37306	86.91589
69	81.6568	52.67176	94	96.90722	88.67925
70	82.35294	53.84615	95	97.4359	90.47619
71	83.04094	55.03876	96	97.95918	92.30769
72	83.72093	56.25	97	98.47716	94.17476
73	84.39306	57.48031	98	98.9899	96.07843
74	85.05747	58.73016	99	99.49749	98.0198
75	85.71429	60	100	100	100

As expected, the graphic representation of the adjustment (Figure 3.4) is symmetrical to the error (Figure 3.3) around the diagonal:

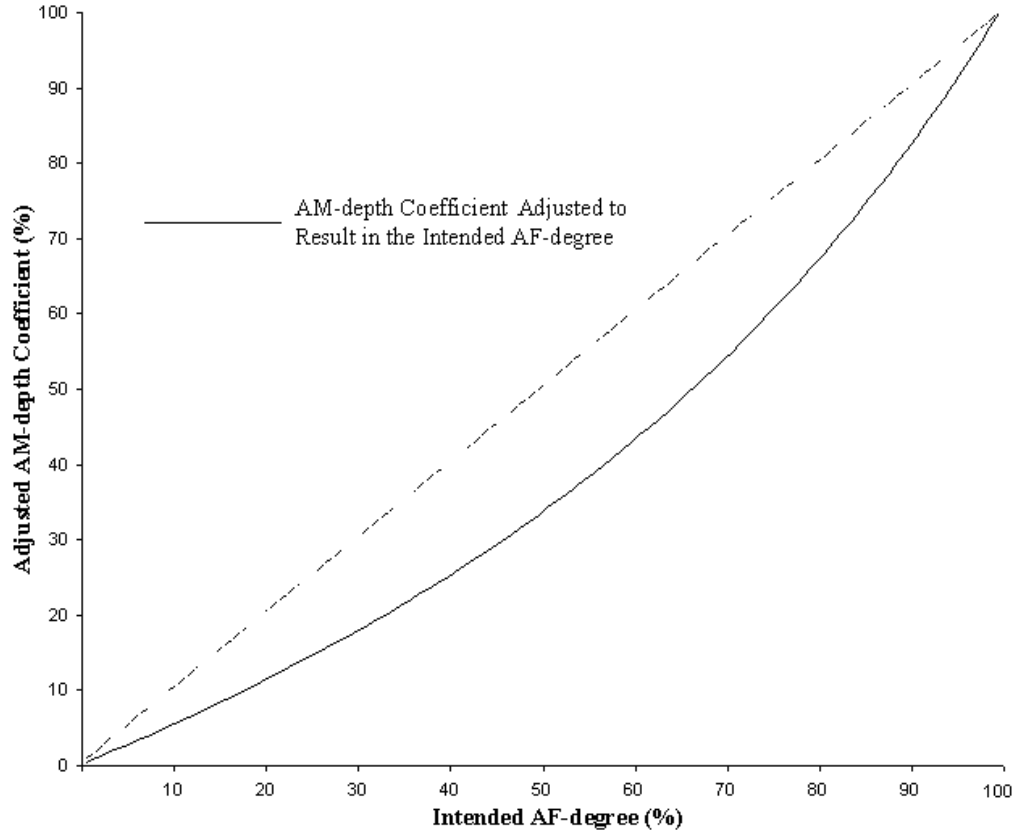


Figure 3.4: Adjustment to the implementation of AF-degree, based on Eq. (3.8). Relationship between intended AF-degree and adjusted AM-depth coefficient that needs to be inserted in Eq. (3.7) for the intended AF-degree to be applied.

The proposed AM-depth coefficient adjustment is necessary for Eq. (3.7) to accurately represent a signal's degree of amplitude fluctuation. The following Section discusses implications of this adjustment to studies that assume AF-degree = AM-depth and examine perceptual correlates of the degree of amplitude fluctuation through manipulation of amplitude modulation depth.

3.4 Implications of the Proposed Adjustment to Perceptual Studies

Manipulating Degree of Amplitude Fluctuation

The implications of the proposed AM-depth coefficient adjustment will be illustrated in a reexamination of a study (Terhardt, 1974a) that is relevant to the present dissertation. In this study, among other issues, Terhardt examined experimentally: a) the influence of modulation depth m and sound pressure level SPL on the roughness of amplitude modulated sines, and b) the relationship between the roughness of amplitude modulated sines (modulation frequency f_{mod} , modulation depth $AM_{depth} = 1$) and the roughness of sine-pairs with amplitude fluctuations of the same rate ($|f_1 - f_2| = f_{mod}$) and degree ($A_1 = A_2$). The reason for reexamining this study is that Terhardt erroneously assumed AF-degree = AM-depth and examined the relationship between degree of amplitude fluctuation and roughness by manipulating amplitude modulation depth. The functions introduced by Terhardt to describe the above relationships will therefore have to be adjusted accordingly. The following discussion does not question Terhardt's experimental method or data. It simply demonstrates that the labels attached to the data misdirected their interpretation.

3.4.1 Influence of Fluctuation Degree on the Roughness of AM Tones

Terhardt (1974a: 203) found that changes in roughness by a factor of 0.5 correspond to changes in AM-depth by an average factor of 0.7 (assuming AF-degree = AM-depth). Data were obtained for various starting AM-depths and carrier frequencies, as well as for both experimental conditions: *half as rough* and *twice as rough*, giving a mean of 0.707 ($SD = 0.0198$). The adjusted mean, calculated from the adjusted data points using Eq. (3.6) (Figure 3.5), is 0.80 ($SD = 0.0225$).

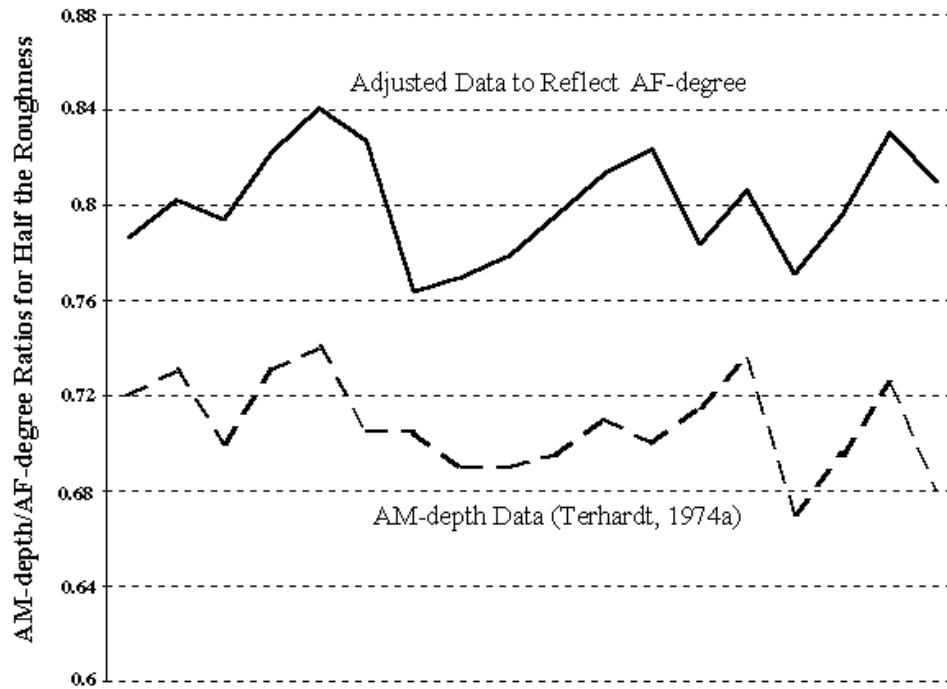


Figure 3.5: AM-depth (after Terhardt, 1974a) and AF-degree (adjusted using Eq. (3.6)) ratios that result in half the roughness.

The power function that describes the relationship between degree of amplitude fluctuation AF_{degree} and degree of perceived roughness R must therefore be changed from $R = cAF_{degree}^2$ to:

$$R = cAF_{degree}^{3.11} \quad (c: \text{constant}) \quad \text{Eq. (3.9)}$$

This change has implications to models estimating the roughness of complex tones (i.e. Helmholtz, 1885; Plomp & Levelt, 1965; Kameoka & Kuriyagawa, 1969a&b; Hutchinson & Knopoff, 1978; Sethares, 1998). It demonstrates that the general assumption of a linear relationship between degree of amplitude fluctuation and the sensation of roughness of AM tones is incorrect and largely underestimates the dependence of roughness on amplitude fluctuation.

3.4.2 Influence of Sound Pressure Level (*SPL*) on the Roughness of AM Tones

Terhardt (1974a: 204) found that, in terms of roughness, reducing the *SPL* of an AM tone ($AM_{depth} = 1$) from 80dB to 40dB (~ 0.0625 factor) is equivalent to reducing the modulation depth of an AM tone (*SPL*: 80dB) from $AM_{depth} = 1$ to $AM_{depth} = 0.83$ (assuming AF-degree = AM-depth). Based on the adjusted data points, the above *SPL* reduction is equivalent to reducing the degree of amplitude fluctuation of an AM tone (*SPL*: 80dB) from $AF_{degree} = 1$ to $AF_{degree} = 0.91$. This corresponds to a reduction in roughness R by a factor of 0.75 (calculated using Eq. (3.9)), instead of the reported 0.69, and can be approximated by the power function:

$$R = cSPL^{\frac{1}{10}} \quad (c: \text{constant}) \quad \text{Eq. (3.10)}$$

Eq. (3.10) indicates that the contribution of *SPL* to the sensation of roughness of AM tones is even smaller than indicated by Terhardt and practically negligible when compared to the contribution of the degree of amplitude fluctuation. Therefore, roughness estimation models that assume roughness to be proportional to the product of interfering sines' amplitudes largely overestimate the influence of *SPL* on roughness.

3.4.3 Roughness Threshold of AM Tones as a Function of Modulation Depth, Fluctuation Degree, and Modulation Frequency

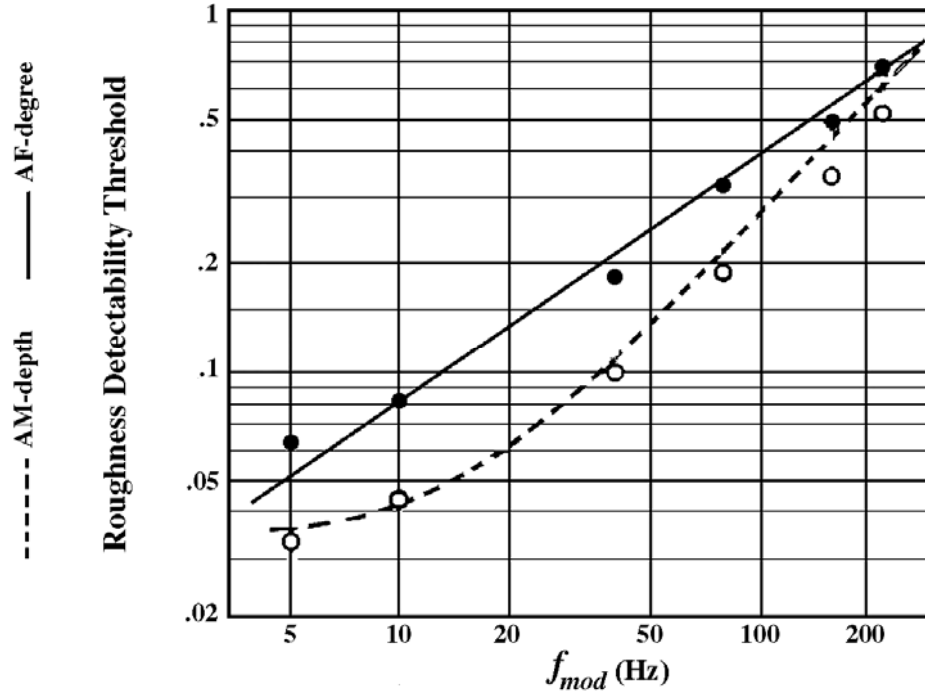


Figure 3.6: Roughness detectability threshold as a function of AM-depth, AF-degree, and modulation frequency (f_{mod}). AM-depth threshold data from Terhardt (1974a: 206) (open circles), approximated by a low-pass RC network (dashed line) with a time constant of 13ms, versus adjusted AF-degree data (filled circles) calculated using Eq. (3.6). Solid line: best linear fit to the adjusted data points.

According to Terhardt (1974a: 207-208), the AM-depth roughness detectability threshold as a function of modulation frequency (Figure 3.6, open circles) is approximated by a low-pass RC network with time constant of 13ms (dashed line),

assuming $AF\text{-degree} = AM\text{-depth}$. Plotting the AF-degree roughness threshold data (adjusted using Eq. (3.6)) reveals a linear relationship (solid line). Terhardt interpreted his data as providing support for the hypothesis of a low-pass weighting mechanism operating on a signal's envelope at a neural level.²² The adjusted data reveals a function with a simpler shape that does not support this hypothesis. Instead, the linear increase of the AF-degree roughness detection threshold with modulation frequency is consistent with the gradual decrease of the roughness sensation with increase in modulation frequency for signals with fixed AF-degree.

²² This hypothesis is part of Terhardt's larger theoretical model on pitch perception (1974b, 1984).

3.4.4 Roughness of a Beating Tone-Pair

Terhardt (1974a: 207-208) reported that the roughness difference between an AM tone (f_{mod} , $AM_{depth} = 1$) and a sine-pair ($|f_1 - f_2| = f_{mod}$; $A_1 = A_2$) corresponds to a reduction in the modulation depth of the AM tone by a factor of 2/3 (assuming AF-degree = AM-depth). Using Eq. (3.6), the same roughness difference corresponds to a reduction in AF-degree by a factor of 4/5. Terhardt interpreted his results as additional evidence supporting the neural low-pass weighting hypothesis discussed in the previous Section. Again, the adjusted results do not support this hypothesis.

Placing the adjusted value, 4/5, in Eq. (3.9) indicates that the roughness R_{AM} of an AM tone (modulation frequency f_{mod} , modulation depth $AM_{depth} = 1$) is related to the roughness R_{pair} of a sine-pair with amplitude fluctuations of the same rate ($|f_1 - f_2| = f_{mod}$) and degree ($A_1 = A_2$) as follows:

$$R_{pair} = 0.5R_{AM} \quad \text{Eq. (3.11)}$$

3.4.5 Discussion

Section 3.4 examined the implications of the error arising when degree of amplitude fluctuation and amplitude modulation depth are assumed to be equivalent. It was shown [Eqs. (3.9), (3.10) & (3.11)] that: a) the importance of AF-degree to the sensation of roughness is larger than indicated by Terhardt (1974a) and b) existing models estimating the roughness of sine-pairs (i.e. Helmholtz, 1885; Plomp & Levelt, 1965; Kameoka & Kuriyagawa, 1969a&b; Hutchinson & Knopoff, 1978, Sethares, 1998) underestimate the contribution of degree of amplitude fluctuation (relative amplitudes values), while overestimating the contribution of *SPL* (absolute amplitude values).²³ It was also shown that the identified error has provided unjustified support to a hypothesis of a neural low-pass weighting mechanism operating on a signal's temporal envelope.

As a general observation, all psychometric functions provided in studies based on the miscalculation of amplitude fluctuation degree have been distorted by a factor proportional to the error described by Eq. (3.6). The influence of the proposed adjustment to the interpretation of distorted psychometric functions will vary according to the meaning attributed by each study to each specific function's shape.

²³ Section 6.4 will introduce a new roughness estimation model that better represents the relationship between the amplitudes of interfering sines and roughness.

3.5 Summary

Chapter 3 identified important qualitative and quantitative differences between *amplitude fluctuation* and *amplitude modulation* and discussed the implications associated with their confusion. It was argued that whenever amplitude modulation depth values (AM-depth) are used as a measure of a modulated signal's degree of amplitude fluctuation (AF-degree), modulation implementation produces an error (Eq. (3.6)) arising from the nonlinear relationship between the presumed and applied AF-degrees. A conceptual and quantitative distinction has been made between AM-depth and AF-degree:

- a) AM-depth refers to the a signal's degree of spectral energy spread and
- b) AF-degree refers to a signal's degree of amplitude fluctuation.

It has been shown that, in order to apply an intended AF-degree, an adjusted coefficient m has to be inserted in the modulation implementation equation, Eq. (3.1). This necessary adjustment (Eq. (3.8)) influences the interpretation of studies where presumed changes in AF-degree have been correlated to changes in physical, physiological or psychological measurements. Therefore, the results and conclusions of studies that manipulate AF-degree through systematic implementation of AM-depth (using Eq. (3.1)) may need to be revised.

In a study by Terhardt (1974a), for example, the identified error has provided unjustified support to a hypothesis of a neural low-pass weighting mechanism operating on a signal's temporal envelope. Additionally, it has supported roughness estimation models of complex spectra that largely underestimate the influence of the degree of amplitude fluctuation (or, in the case of pairs of sines, of relative amplitude values) to the roughness of complex tones. This will be addressed further in the following chapters. Studies that have used a fixed AM-depth value of 100% (i.e. de Boer, 1956a&b; Schouten et al., 1962; Schroeder, 1966; Iwamiya, 1995; etc.) are not affected by the adjustment. This is also the case with studies using any fixed AM-depth value, providing it was used to produce a desired spectrum (i.e. Smoorenburg, 1970) and not to represent a specific degree of amplitude fluctuation (i.e. Ritsma, 1962, 1963; van den Brink, 1970; Terhardt, 1974a; Iwamiya & Fujiwara, 1985; Iwamiya & Okamoto, 1996; etc.).

In conclusion, the erroneous assumption of the equivalence between amplitude modulation and amplitude fluctuation has resulted in the misinterpretation of studies examining perceptual correlates of amplitude fluctuation, providing questionable support to related perceptual models, and distorting the resulting psychometric functions by a factor proportional to the error (Eq. (3.6)). The present dissertation will keep the terms *amplitude modulation* and *amplitude fluctuation* separate, adopting the term *amplitude fluctuation* to refer specifically to variations of a signal's amplitude around a reference value.

Chapter 4 Amplitude Fluctuation and Sound:

Background, Definitions, and Clarification of Terms

4.1 Wave Processes and Sound Waves in Air

Agreeing on a single, all-encompassing definition for a wave is non-trivial. A vibration can be defined as a back-and-forth motion around a point of rest (i.e. Campbell & Greated, 1987: 5) or, more generally, as a variation of any physical property of a system around a reference value. However, defining the necessary and sufficient characteristics that qualify a phenomenon to be called a wave is, at least, flexible. The term is often understood intuitively as the transport of disturbances in space, not associated with motion of the medium occupying this space as a whole. In a wave, the energy of a vibration is moving away from the source in the form of a disturbance within the surrounding medium (Hall, 1980: 8). However, this notion is problematic for a standing wave (i.e. a wave on a string), where energy is being transformed rather than moving, or for electromagnetic / light waves in a vacuum, where the concept of medium does not apply.

For such reasons, wave theory represents a peculiar branch of physics that is concerned with the properties of wave processes independently from their physical origin. The peculiarity lays in the fact that this independence from physical origin is accompanied by a heavy reliance on origin when describing any specific instance of a wave process. For example, acoustics is distinguished from optics in that sound waves are related to a mechanical rather than an electromagnetic wave-like transfer / transformation of vibratory energy. Concepts such as mass, momentum, inertia, or elasticity, become therefore crucial in describing acoustic (as opposed to optic) wave processes. This difference in origin introduces certain wave characteristics particular to the properties of the medium involved (i.e. in the case of air: vortices, radiation pressure, shock waves, etc., in the case of solids: Rayleigh waves, dispersion, etc., and so on).

Other properties, however, although they are usually described in an origin-specific manner, may be generalized to all waves. For example, based on the mechanical origin of acoustic waves there can be a moving disturbance in space-time if and only if the medium involved is neither infinitely stiff nor infinitely pliable. If all the parts making up a medium were rigidly bound, then they would all vibrate as one, with no delay in the transmission of the vibration and therefore no wave motion (or rather infinitely fast wave motion). On the other hand, if all the parts were independent, then there would not be any transmission of the vibration and again, no wave motion (or rather infinitely slow wave motion). Although the above statements are meaningless in the case of waves that do not require a medium, they reveal a

characteristic that is relevant to all waves regardless of origin: within a wave, the phase of a vibration (i.e. its position within the vibration cycle) is different for adjacent points in space because the vibration reaches these points at different times.

Similarly, wave processes revealed from the study of wave phenomena with origins different from that of sound waves can be equally significant to the understanding of sound phenomena. An example relevant to the present study is Young's principle of interference (Young, 1802, in Hunt, 1978: 132). This principle²⁴ was first introduced in Young's study of light and, within some specific contexts (i.e. scattering of sound by sound), is still a researched area in the study of sound. As another example, the phenomenon of dispersion²⁵ demonstrates that wave modulations behave as regular waves. When modulations propagate in media where the speed of wave propagation depends on frequency, they separate from the complex wave they belonged to and travel independently carrying energy, similarly to the rest of the frequency components of the complex wave. It is true that this separation will never happen in a non-dispersive medium such as air, where all frequencies move with the same speed. Nonetheless, the important point is that the dispersive case serves to illustrate that modulations in general and amplitude fluctuations in particular behave as waves.

²⁴ The principle of interference is related to the principle of superposition and states that the combined amplitude of two or more waves may be larger (constructive interference) or smaller (destructive interference) than the amplitude of the individual waves depending on their phase relationship.

²⁵ The term *dispersion* describes a dependence of wave velocity on frequency.

Although the air is a non-dispersive medium, the dispersive case is addressed in this study for two related reasons:

- a) Dispersion provides a case where modulations are isolated from the waves that carry them and can therefore be studied easier (assuming that the only characteristic that changes during dispersion is the modulations' velocity).
- b) Systems with dispersion provide better cases for the mathematical analysis of the kinematic properties of waves (i.e. frequency, wavelength, phase and group velocities). Such an analysis is crucial to the study of modulations (Ostrovsky & Potapov, 1999: 1), of which amplitude fluctuations are just a special case.

Introducing dispersion theory to the study of musical sound is not as unusual as it may first appear. The fact that the study of dispersion theory usually starts by examining the propagation of energy concentrations in the phenomenon of beats indicates that linking dispersion to modulations/amplitude fluctuations is commonplace rather than the exception. With energy concentrated in two distinct frequency regions (modulation rate and carrier frequency), the phenomenon of beats provides a good example of a situation that exhibits the difference between a wave's group and phase velocities²⁶ (Towne, 1967: 379). At the same time, the difference between group and phase velocities makes it possible to observe the energy content of beats at work. In music, sound diffraction, absorption, and reverberation are examples

²⁶ The concepts of group and phase velocities are addressed in Section 5.2.2.

of phenomena that have been better understood with the aid of dispersion theory.²⁷

The present study will illustrate the usefulness of dispersion theory to the understanding of amplitude fluctuations as waves.

In conclusion, the term *wave* implies three general notions: vibrations in time, disturbances in space, and moving disturbances in space-time associated with the transfer/transformation of energy. Based on these notions, the following origin-specific definition is proposed for sound waves in air:

"Sound-waves in air represent a transfer of vibratory energy characterized by: i) rate (frequency), ii) starting position (phase), and iii) magnitude (amplitude) of vibration. In general, amplitude can be expressed equivalently in terms of maximum displacement, velocity, or pressure relative to a reference value. Sound waves in air are manifested as alternating air-condensations and rarefactions that spread away from the vibrating source with a velocity usually not related to the velocity amplitude of the vibration. They result in pressure/density disturbance patterns in the surrounding medium, which, in general, correspond to the signal that plots the vibration of the source over time."

The proposed definition will serve as an initial operational definition of sound waves in air to which further qualifications may be added as needed.

²⁷ Diffraction, absorption, and reverberation are frequency dependent phenomena so dispersion theory is better fit to address them. Modeling all three phenomena requires the introduction of a frequency-dependent coefficient that is an analogue to the dispersion relation characterizing a dispersive medium (Towne, 1967: 196-197, 221-223).

4.2 Beating as a Sensation

The most familiar perceptual manifestation of amplitude fluctuation is the phenomenon of beats. This phenomenon was addressed relatively early by scientists such as Leonardo da Vinci (early 1500s) and Huyghens (late 1600s) in their exposition of the wave superposition principle (in Hunt, 1978: 76), and by Beeckman (1620s, in Plomp, 1966: 462) in his work on tuning. The theoretical treatment of beats, however, began a continuous tradition only after the work of Sauveur (early 1700s, in Carlton, 1990). Since then, the term *beating* has been used in the music cognition and acoustics literature to describe both a perceptual and a physical phenomenon. In the perceptual frame of reference, it describes a slow loudness fluctuation that results when listening to two tones slightly out of tune. In the physical frame of reference, it describes the slow amplitude fluctuation of a vibration / wave resulting from the superposition of two vibrations / waves with nearly equal frequencies. The double use of any term can be confusing and, in the case of beating, is also questionable for a number of reasons:

- Historically, the term was first introduced by musicians to describe the sensation that arises when listening to two, slightly out of tune, musical tones. This historical precedence is alluded to in the following excerpt from *Histoire de*

l'Academie Royale des Sciences (1700), as a comment on Sauveur's attempts to theoretically address the phenomenon:

It seems, indeed, that the common expression of musicians, who say that the [organ] pipes “beat” when their sound is reinforced in this way [i.e. by doubling one pipe with a second one slightly out of tune], has its origin in this [i.e. Sauveur's] idea. (In Carlton, 1990: 18).

Sauveur's idea of the physical phenomenon behind the beating sensation was quite different from our current understanding, reflecting an equally different idea of sound waves that will be discussed in Section 4.3. The term *beating* was introduced probably to point to the similarity between the sound of two tones slightly out of tune, whose loudness slowly fluctuates, and the sound of a beating heart.

- In the physical frame of reference, there seems to be no reason why the phenomenon has to be limited to low frequency amplitude fluctuations that result from the superposition of vibrations with nearly equal frequencies (qualifications that accompany the term *beating*). Whenever two vibrations with different frequencies are added together, the amplitude of the resulting vibration will fluctuate. The rate and degree of fluctuation can be calculated / predicted by the amplitude and frequency values of the original vibrations, regardless of their frequency or amplitude difference. The only condition is that the average frequency is much larger than the frequency difference. If this condition is not satisfied (as is the case with the beats of mistuned consonances, i.e. the mistuned

fifth: 200Hz - 301Hz), the calculation method may be different, but the perceptual attributes of the rate and degree of amplitude fluctuation remain the same. The fact that the beating sensation disappears beyond a certain rate / degree of fluctuation does not render the physical phenomenon (amplitude fluctuation) any less continuous, and understanding it as such can be misleading.

- Vibrations, waves, and the sensation of sound are not isomorphic. Although this is well known, especially with regards to waves and the sensation of sound, the way the sensation of beating is approached reflects the questionable assumption that perceptual changes, discontinuities, or categories, should necessarily correspond to physical ones. At the same time, vibrations and waves are being treated as equivalent, fact indicated by the interchangeable use of the terms. Observations on vibrations are expected to hold as they are for waves and observations on waves are expected to have a direct source on vibrations, expectations that have been questioned since Rayleigh's *The Theory of Sound* (1896).

Based on the above arguments, the present study will reserve the term *beating* or *beats* to describe the perceptual phenomenon of loudness fluctuations.

4.3 Sound Waves and Beats

The idea of the compressing *and* expanding action of air on air, fundamental to the understanding of air-borne sound waves, was obscured for centuries by the insistence on the abused (Hunt, 1978: 23) analogy between surface waves in water and sound waves in air.²⁸ The notion of air compression was first alluded to in Aristotle's (ca. 384-322 B.C.) *De Anima* (in Hunt, 1978: 23) and explicitly stated in *De Audibilibus*, a manuscript attributed by some (i.e. Pierce, 1989: 2) to Aristotle and by others (i.e. Hunt, 1978: 26) to Straton of Lampsacus (ca. 340-269 B.C.):

[Sound propagates] not due to the air taking on a shape as some thing [i.e. water], but ... by contraction, expansion, and compression. ... For when the breath that falls on it strikes the air with successive blows, ... [the air thrusts forward] in like manner the adjoining air, so that the sound travels unaltered in quality as far as the disturbance of the air manages to reach. (In Hunt, 1978: 26).

²⁸ The analogy between the propagation of waves in water and of sound in air was first formally noticed by Vitruvius, a Roman engineer and architect, around the middle of the first century B.C. (in Campbell & Greated, 1987: 23). It was not fully expressed until the sixteenth century A.D. by Leonardo da Vinci who first noted that the motion of particles of the medium and the propagation of waves are different processes (Ostrovsky & Potapov, 1999: 15). Pointing out the analogy between waves in water and sound in air was therefore very useful in developing an understanding of sound propagation in air. It was the insistence on this analogy for all aspects of sound propagation that hindered the appreciation of the differences between transverse and longitudinal waves in air and of the role of the pressure variations associated with the latter.

By the time of Newton (late 1600s), however, a firm belief had been established²⁹ that sound propagates through the air compressions, with only the condensed portion of a medium being able to communicate its motion to the rest of the medium or to a boundary (i.e. eardrum). Sound waves were understood as successions of short pressure impulses. Each impulse was assumed to be independent from the other impulses in the succession and unaffected by any other such succession, unless there was a precise impulse coincidence.³⁰ Attention was focusing on one half of the whole process, ignoring the potential energy of the rarefied portions of the medium³¹ and preventing the recognition of the possibility of interference. Newton's work towards a theory of tides expressly discusses the possibility of the cancellation of water waves if the ridge of one wave would coincide with the valley of another. Although this (in retrospect) points to nothing less than the possibility of destructive interference, it did not attract attention probably because of the highly context-specific

²⁹ This belief is reflected in Newton's mathematical model of sound propagation (1686; qualitative analysis in Carlton, 1990: 7-9), which included a mechanical interpretation of sound as being a series of pressure pulses transmitted through neighboring fluid particles.

³⁰ Based on this approach, two tones in Unison would not sound louder than a single tone unless the two tones started at times and positions appropriate for their impulses to exactly coincide (expectation that contradicts experience). It can only be assumed that this logical path was never followed, or that any pulse-train addition was supposed to be happening *after* the sound had already entered the ear. Such an assumption has some parallels to the contemporary assumption that wave interference products are subjective, caused just by the nonlinear response of the ear, or by some additional, higher level interaction of the sound information.

³¹ Aderald (12th century) pointed out that a negative-pressure wave (rarefaction) may be transmitted in the same way as an impulse (condensation). "Just as when I form a sound by impelling the air from me, ... so when I draw the air inwards, and withdraw it violently from others, it is ... perceived." (In Hunt, 1978: 62). This correct assessment was followed, however, by a wrong conclusion linking the phenomenon now known as diffraction to the alleged porous nature of all solids

nature of its presentation. Young referenced Newton's work during his heated arguments with Cough³² against claims of the latter that the difference tone (to be discussed later) is subjective, created in the listener's mind. In a letter to Cough, Young drew an analogy between the difference tone and a case of compounded tidal motions: *"If any person should insist that the phenomenon is not a tide, but just an alternate elevation and depression of the water, and that it exists only in the sensations of the observers, and not in the sea, I should be very little disposed to enter into arguments with him on the subject."* (Young, 1803, in Carlton, 1990: 87-88).

It was on the incorrect understanding of sound waves as pressure impulses that Sauveur based his theoretical explanation of beats. For Sauveur, sound waves were pulse-trains, and beats were a manifestation of the precise coincidence of the pulses of two pulse-trains with different frequencies. For example, if a sound resulting from $f_1 = 100$ pulses per second³³ was combined with a sound resulting from $f_2 = 102$ pulses per second, their pulses would precisely coincide only $f_2 - f_1 = 2$ times per second,

that was supposed to allow the passage of air and therefore sound through them, hurting the credibility of the entire statement.

³² Cough was a botanist and amateur physicist from Manchester. His arguments with Young on sound interference started when Young criticized Smith's explanation of beats, and became heated after Cough published his "Theory of Compound Sounds" (1802, in Carlton 1990: 70). Cough's theory agreed with Smith and reflected the general resistance of the scientific community to Young's principle of interference. Due to the popularity and controversial nature of the topic, the dispute between Young and Cough was conducted in public through the pages of *Annalen der Physik* (in Carlton, 1990: 65).

³³ Sauveur used frequency ratios rather than absolute frequency values for his calculations and did not explicitly state the relationship between the beating rate and the frequencies of the beating tones. His theory of beats actually arose as an aside in his search for a reproducible pitch (or rather frequency) standard. In this study he used as a reference the beating rate of a just Minor 2nd, an interval arrived at

giving rise to a beat sensation with that rate. In such an interpretation, the beating sensation corresponds to a specific pulse-train and in turn to a sound wave, sharing all characteristics of waves, including the fact that it carries energy.

Sauveur's theory of beats was very similar to Galileo's theory of consonance, which was also based on the degree of impulse coincidence (Carlton, 1990).

According to this theory, the more frequent the coincidences of impulses the higher the degree of consonance, a conclusion that openly contradicted observation. For example, a slightly mistuned Octave should, according to this theory, be much more dissonant than a Minor 2nd. In an effort to alleviate this problem, Smith (mid 1700s) developed a different theory of beats (in Carlton, 1990: 24-35), which was essentially a complex and forced attempt to fit observation into the problematic understanding of sound waves as pulse-trains.³⁴

The confusion was increased further when DeMorgan (1864, in Carlton, 1990: 34) tried to justify both, Sauveur and Smith, by claiming that they were describing two different kinds of beats: one between mistuned Unisons and one between mistuned consonances (i.e. fourths, fifths, etc. with frequency ratio slightly removed from a small integer ratio). DeMorgan based his claim on an observation by Smith, in which

in a stepwise manner from larger intervals whose justness was assumed by the absence of any beating (i.e. beating that would have resulted from the mistuning of consonances).

³⁴ In his mathematical formulations, Smith did manage to prove that the beating frequency is equal to the difference of the generating tones' frequencies. This conclusion was rediscovered by Hällstrom (1819, in Carlton, 1990: 30) based on the correct physical reasoning, and was widely circulated after his second and revised publication on the subject (1832).

'just' consonances gave rise to a 'fluttering' sound upon which the beating was added when the consonances were detuned. While the detuning was influencing the beating rate it did not seem to have any noticeable effect on the 'fluttering' sound.

Based on the modern understanding of wave superposition and interference, Smith and DeMorgan's observations do not indicate the presence of two different kinds of beats, but rather different manifestations of the same phenomenon. For justly tuned consonances between complex tones, the frequencies of certain mid and high components coincide. As the consonances are detuned, the frequencies of these components start to shift, resulting in slow amplitude fluctuations that give rise to the beating sensation. On the other hand, the frequencies of the fundamentals are always different and result in amplitude fluctuations that may give rise to the sensation of 'fluttering' or roughness depending on the rate of fluctuation. The slight detuning of the consonances does not affect this sensation, which persists for a relatively wide range of frequency differences.

The only problem with the above explanation is that beating sensations have also been observed in the case of mistuned consonances between sines (i.e. Helmholtz, 1885: 197-211; Wegel & Lane, 1924, in Plomp, 1966: 463; von Békésy, 1960: 577-590; Plomp, 1966). Early explanations of this phenomenon were based on the nonlinear creation of combination tones (Helmholtz, 1885: 197-211) or harmonics (Wegel & Lane, 1924, in Plomp, 1966: 463) inside the ear.

However, in his arguments against using the *best beats* method to detect aurally created harmonics of a sine tone,³⁵ von Békésy demonstrated that mistuned consonances of sines do give rise to beating-like sensations even when applied to the skin of the arm (1960: 579-582). This observation challenged the early explanations leaving open the question of beating in the case of mistuned consonances between sines. From within the context of the present study, the signals of mistuned consonances of sines exhibit amplitude fluctuations at the observed beating rates and degrees (see the Appendix) and can therefore explain them without the need for a nonlinear generating mechanism, whether at a cochlear or a neural level.

Returning to the development of the understanding of sound waves and the sensation of beating, the available theoretical explanations did not improve until the exploration of the phenomena of diffraction and interference (within the context of a theory of light) laid the foundations for contemporary wave theory. It is interesting to note that, although most of the theoretical advancements in optics were arrived at by means of drawing analogies between light and sound, the same advancements were not applied to the theory of sound until later. Grimaldi's publication on the bending of

³⁵ The *best beats* method can be described as follows: In order to examine the existence of aurally created harmonics of a sinusoidal stimulus, another sine is presented with a frequency slightly removed from that of the hypothetical aural harmonic. The observation of beats is considered as evidence for the presence of the specific aural harmonic. The amplitude of the second sine for which the beats sound strongest (or *best*) is considered as equivalent to the amplitude of the hypothetical aural harmonic since interference of two sines with equal amplitudes results in the strongest beats.

light around corners (1665, in Hunt, 1978: 131) depended heavily on his observation of a similar bending of sound, while Young (1802, in Hunt, 1978: 132) introduced the principle of interference to the theory of light based on the similarity between optical interference and the phenomenon of beats, before the phenomenon of beats had been theoretically addressed. Diffraction and interference became the central issues in the dispute between the wave and corpuscular theories of light, and Newton's insistence on the corpuscular viewpoint resulted in the suppression of the wave theory. Fresnel (early 1800s) reinforced the wave viewpoint by offering a mathematical analysis of both interference and diffraction based on Huygens's principle (Huygens, 1678; in Hunt, 1978: 133). This principle states that successive portions of a wave-front are determined by the envelope of secondary wavelets. It is widely used in the study of wave propagation and is included in the present study's argument for the wave nature of amplitude fluctuations.

It was only during the beginning of the nineteenth century that continuum physics, or field theory, received a mathematical structure, with the general wave equation emerging in a number of contexts including the propagation of sound-waves in air.³⁶ The earlier 'quantal' understanding of sound waves as pulse-trains was

³⁶ A theory of sound propagation based on firm mathematical and physical concepts was initiated during the 18th century by Euler (1707-1783), Lagrange (1736-1813), and d'Alembert (1717-1783). Modern theories can be regarded for the most part as refinements on that theory. The general wave equation, also called *d'Alembertian*, is based on Euler's equation of motion for a fluid and is outlined in Pierce (1998: 27).

replaced by the 'analogue' conception of sound waves in air as continuous pressure/density variations. The definition of sound waves offered in Section 4.1 has been based on this conception.

From within such an understanding, the sensation of beating does not correspond to a pulse train made of coincident pulses belonging to other pulse trains, but to a periodic and gradual alteration between constructive and destructive interference of the interacting sound waves, resulting in signal amplitude fluctuations.

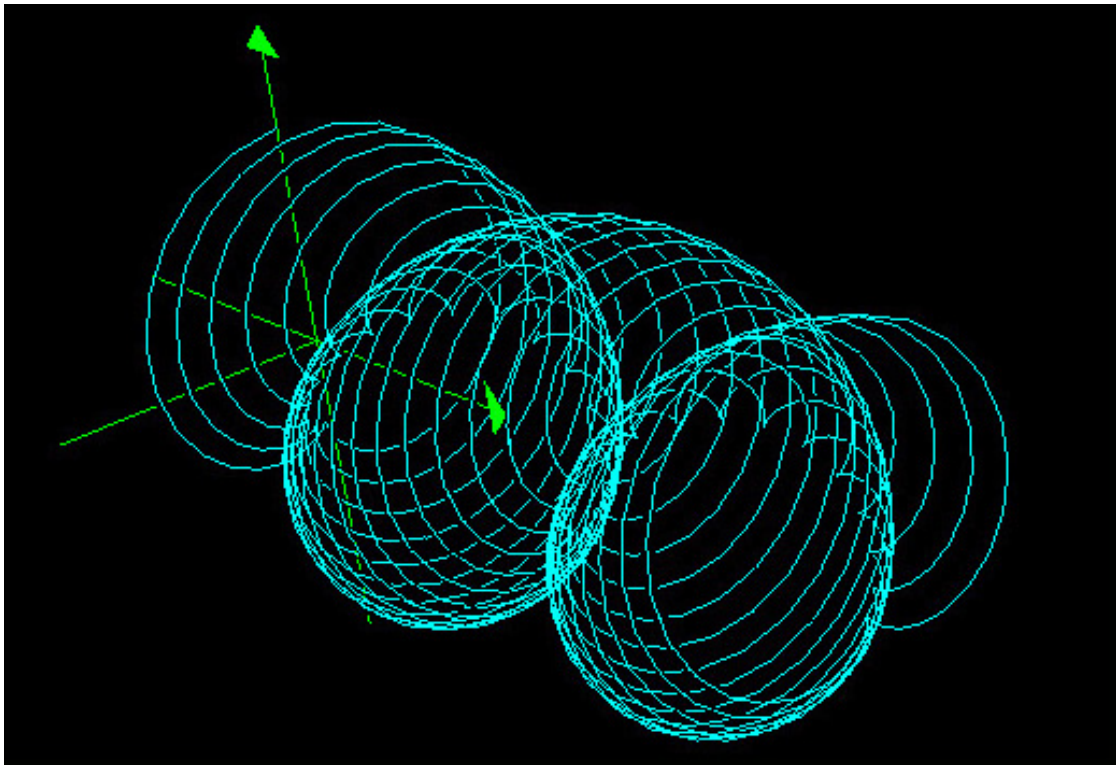


Figure 4.1: Three-dimensional signal of a mistuned octave between sines: $f_1=30\text{Hz}$, $f_2=62\text{Hz}$, $A_1=1$, $A_2=0.5$. Three-dimensional twisted spiral representations of complex waves and their relationship to traditional two-dimensional signals are addressed in Sections 5.2.1 & 6.3.2 and in the Appendix.

Chapter 5 Perceptual Manifestations of Amplitude

Fluctuation: Problem Analysis, Suggestions

5.1 Beats and Amplitude Fluctuations

The connection between the sensation of beating and amplitude fluctuations has been established since the works of Helmholtz (1885) and Rayleigh (1896). However, the insistence on considering beating a manifestation of a linear phenomenon suggests that little attention has been placed on the discrepancy between the beating rate and the modulation rate indicated by the mathematical expression of interference.³⁷ This discrepancy is usually attributed to the hearing mechanism, although there has been enough physical evidence indicating that actual energy is carried at a rate double the mathematically determined modulation rate:³⁸

Beats³⁹ ... exist objectively as disturbances in the air before it [the air] reaches the ear. They are reinforced by resonators, disturb sand etc. In the case of the

³⁷ For two interfering sines, the computed modulation rate is half the beating rate.

³⁸ Such evidence is presented in the present Section and in the discussion on the difference tone (Section 5.4).

³⁹ The term *beats* is used here to refer to amplitude fluctuations.

beats of harmonium reeds in Appunn's tonometer, they strongly shook the box containing the reeds (Helmholtz, 1885: 531).

The principle of linear superposition, first introduced by Bernoulli (late 1700s), came out of the study of strings, considered as one-dimensional systems. It states that many sinusoidal vibrations can co-exist on a string independently of one another and that the total effect at any point is the algebraic sum of the individual motions (Bernoulli, 1755, in Beyer, 1999: 14).⁴⁰ Objections against the idea of amplitude fluctuations carrying energy have occasionally been based on this principle, a principle intended for vibrations in one dimension and not waves in three.

Linearity of superposition applies only to linear combinations of field variables (i.e. velocity) and not to energy quantities such as intensity or power, which are quadratic with respect to field variables. Intensities can be linearly added only in special cases such as the case of two waves traveling in opposite directions (Towne, 1967: 73).⁴¹

⁴⁰ If the compound vibration at any point of a string is the algebraic sum of the individual vibrations, it is the sinusoidal waves (and not the vibrations) that can co-exist on a string independently of one-another. The way the principle of superposition was first stated reflects the problematic assumption of equivalence between vibrations and waves addressed in Section 4.3.

⁴¹ At the same time, one of the conditions for observing the manifestations of amplitude fluctuation is that the interfering waves are collinear (i.e. they travel in the same direction; see Section 5.4.1). The theoretical estimation of the energy content of amplitude fluctuation (Section 6.3.1) will be based partly on the combination of these two statements.

For example, the intensity I of a sound wave with pressure P and velocity V is: $I = PV$. Adding two waves $(P_1, V_1; P_2, V_2)$ that travel in the same direction will result in a complex wave with pressure $P = P_1 + P_2$, velocity $V = V_1 + V_2$, and intensity:

$$I = (P_1 + P_2)(V_1 + V_2) = (P_1V_1 + P_2V_2) + (P_1V_2 + P_2V_1) = I_1 + I_2 + (P_1V_2 + P_2V_1) \quad \text{Eq. (5.1)}$$

Eq. (5.1) demonstrates that $I \neq I_1 + I_2$.

The following Section identifies the limitations of two-dimensional signals in representing sound waves, addressing graphically the problem of linear addition of wave-energy quantities.

5.2 Sound Signals and Energy Content of Amplitude Fluctuations

5.2.1 Drawbacks of Two-Dimensional Signal Representations

The problems arising from considering one-dimensional vibrations equivalent to the resulting three-dimensional waves can be illustrated by the inability of two-dimensional vibration signals (time included) to representing sine-wave energy and amplitude consistently.

5.2.1.1 Two-Dimensional Signals and Energy

Two-dimensional signals with amplitude fluctuation derive from the linear addition of two-dimensional sine signals. In the simplest case, this would be a signal resulting from the addition of two sines with slightly different frequencies. In order for such an addition to be meaningful, variables represented by each signal must share the same units. For two-dimensional signals these variables are:

- Time, which is related to the period and therefore frequency of a signal; usually on the x (horizontal) axis.
- A variable such as displacement, voltage, pressure, etc., whose value oscillates over time around a reference value; usually on the y (vertical) axis.
- The energy drawn/radiated/carried in the form of a wave, whose value is represented by the area outlined by the signal and the x axis.

Even in the case of the simplest two-dimensional signals (i.e. sine signals), only the first two variables are represented with consistent units across frequencies. If two sine signals with different frequencies share the same coordinate system, their relationship will be represented correctly in the case of time and displacement / pressure, but incorrectly in the case of energy. Figure 5.1 illustrates this problem for a two-component complex signal: $f_1 = 10\text{Hz}$, $f_2 = 12\text{Hz}$, $A_1 = A_2$, $\phi_1 = \phi_2 = 0$ rads.

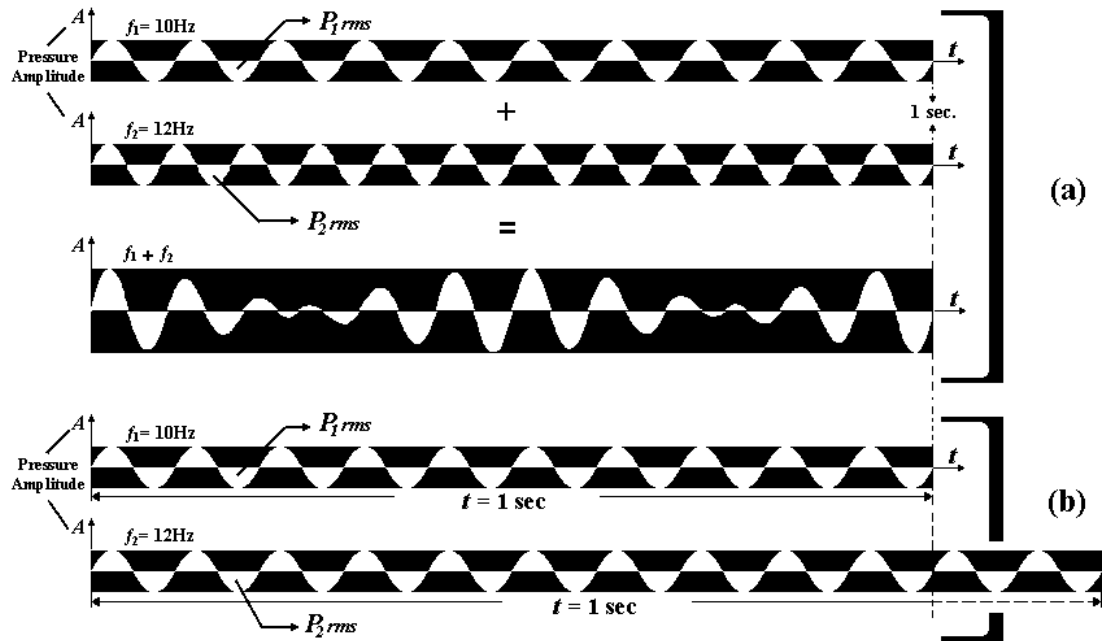


Figure 5.1: Two one-second-long sine signals ($f_1=10\text{Hz}$, $f_2=12\text{Hz}$, $A_1=A_2$), illustrating the problem of energy representation. (a): The time-axis scale is consistent for both signals, suggesting that linear addition might be possible. However, their relative root-mean-square pressures P_{rms} (highlighted areas of f_1 and f_2) have been misrepresented (see text for details). (b): The P_{rms} relationship between the two signals is correct. However, the time-axis scale is inconsistent.

If wave signals are supposed to represent the transfer/transformation of energy within a medium, comparing (or adding, subtracting, etc.) two or more wave signals graphically is meaningful only if the dimension of energy follows a consistent scaling scheme across signals, as is the case with time and displacement/pressure/etc. Therefore, comparing two-dimensional representations (i.e. sinusoidal vibration signals of displacement/pressure/etc. over time) of three-dimensional concepts (i.e. sinusoidal sound waves) is possible only if the collapsed dimension (i.e. energy) also follows a consistent scaling scheme for both representations. In Figure 5.1, the two axes represent pressure and time while the highlighted areas are a measure of the root

mean square pressure (P_{rms}) and therefore of the energy in each signal. When time and pressure follow consistent scaling schemes across signals (i.e. Figure 5.1.a) the representation of the relative P_{rms} values is incorrect. In Figure 5.1.a, $P_{2,rms} = P_{1,rms}$ although the correct relationship is $P_{2,rms} = 1.2P_{1,rms}$.⁴² The two representations follow different scaling schemes for the dimension of energy, questioning the validity of their linear addition. Figure 5.1.b shows the correct relationship between P_{rms} values. However, the time axis for the second signal has been stretched relative to that of the first, forbidding the linear addition of the two signals.

These observations lead to the question: How can one explain that, while a two-dimensional modulated curve derives geometrically from the linear addition of sine curves, the variables it is assumed to represent are distorted in the addition? This is part of a larger representational problem, not limited to the graphic representation of vibrations and waves. It is the problem of assuming that the graphical representation of a concept possesses all of the concept's properties and can therefore replace it and be manipulated independently, giving new representations that are as meaningful conceptually as the original representation was. This assumption is an extension of a more general assumption that equates observations to the concepts invented to describe them and is behind the representational inconsistencies identified in this chapter.

⁴² Root-mean-square of the pressure is proportional to frequency.

5.2.1.2 Two-Dimensional Signals and Amplitude

As mentioned in Chapter 3, sinusoidal amplitude modulation of a sine signal is a nonlinear process that alters the resulting signal's spectral composition. It introduces two sidebands at frequencies determined by the modulation frequency f_{mod} and amplitudes determined by the modulation depth m . More specifically, modulation depth m indicates that the sum of the sidebands' amplitudes will be equal to m multiplied by the amplitude of the original sine signal (carrier). Eq. (5.2) restates the above definition:

$$(A + Am \cos 2\pi f_{\text{mod}} t) \sin(2\pi f t + \phi) = A \frac{m}{2} \sin[2\pi(f - f_{\text{mod}})t + \phi_1] + A \sin(2\pi f t + \phi) + A \frac{m}{2} \sin[2\pi(f + f_{\text{mod}})t + \phi_2] \quad \text{Eq. (5.2)}$$

where f_{mod} and m ($0 \leq m \leq \infty$) are the modulation parameters, and f , A , and ϕ are the parameters of the sine with $|\phi_1 - \phi| = |\phi_2 - \phi| = \pi$ rads ($\pi = 3.14$; t : time).

Amplitude modulation (AM) depth values larger than 100% are possible and have been used in various psychoacoustical studies (i.e. van den Brink, 1970). AM-depth can be infinitely large. Infinite AM-depth indicates that the energy of the side

bands is infinitely larger than the energy of the carrier (equivalent to the carrier having zero energy), resulting in an AM tone with two rather than three components (i.e. Figure 5.3.j). As shown in Chapter 3,⁴³ $AM_{depth} = \infty$ is equivalent to degree of amplitude fluctuation $AF_{degree} = 2$ (200%). Additionally, Eq. (3.2) shows that the minimum amplitude A_{min} of an amplitude modulated sine can only range between the values A_{max} and $-A_{max}$ (A_{max} : maximum amplitude of an amplitude modulated sine). In fact, if $A_{max} = 1$, then $A_{min} = 1 - AF_{degree}$, for any AF-degree value. For example, the signal and spectrum in Figure 5.2 correspond to $AM_{depth} = 2$ (200%), since the energy of the side bands is equal to the energy of the carrier multiplied by 2. Based on Eq. (3.6), this is equivalent to $AF_{degree} = 1.3333$ (133.33%). Therefore, if $A_{max} = 1$ then $A_{min} = 1 - AF_{degree} = -0.3333$

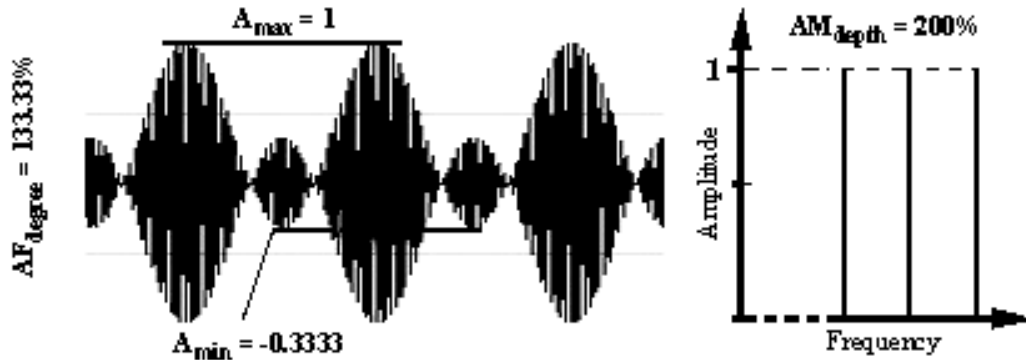


Figure 5.2: Signal and spectrum of a modulated sine with AM-depth = 2 (200%) resulting in AF-degree = 1.3333 (133.33%). AF-degree value derived from Eq. (3.6).

⁴³ Chapter 3 examined the difference between amplitude modulation depth and degree of amplitude fluctuation.

More generally, Eqs. (3.4) & (3.2) demonstrate that AM-depth / AF-degree values larger than 100% always result in negative A_{\min} values. For example, AM-depth = 3 (300%) is equivalent to AF-degree = 1.5 (150%). Placing these values to Eqs. (3.4) & (3.2) respectively and assuming $A_{\max} = 1$, results in $A_{\min} = -0.5$.

Negative amplitude values are not, in themselves, problematic. Whether amplitude values will be positive or negative is a matter of convention. It is a matter of deciding which half of the y-axis in the two-dimensional representation will be labeled positive and which will be labeled negative. The only requirement is consistency of representation so that, within the same signal, labels do not change with time. Combining this requirement with the definition of amplitude as maximum displacement indicates that the same signal cannot have both positive and negative amplitude values. Nonetheless, AM-depth / AF-degree values larger than 100% result in signals with amplitude values that alternate between positive and negative. As can be seen in Figure 5.3, if AM-depth / AF-degree > 100% then A_{\max} is located at the top half of the signal (positive) while A_{\min} is located at the bottom half (negative):

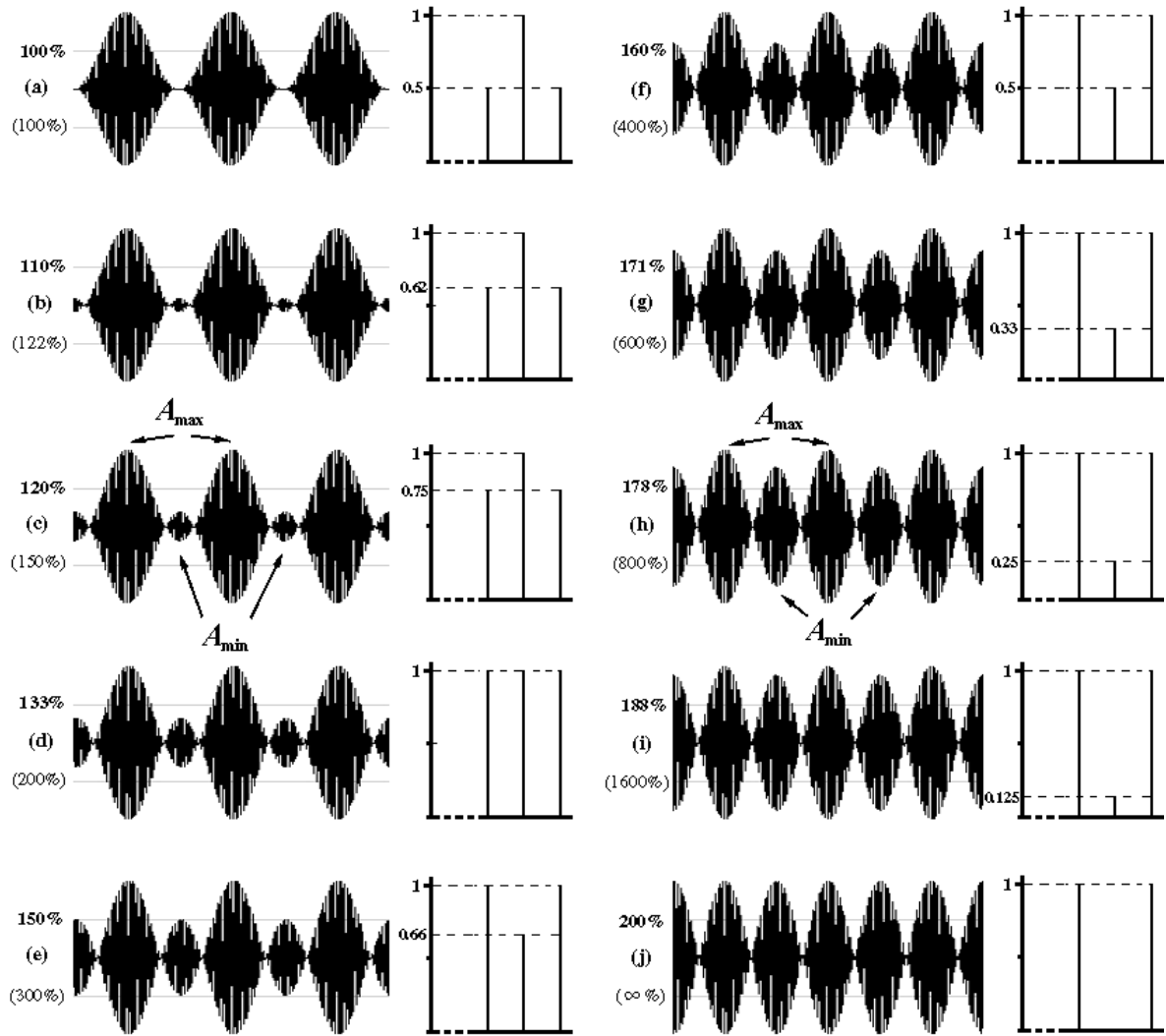


Figure 5.3: Signals (normalized) and spectra resulting from gradual increase in AF-degree of an amplitude-modulated sine from 100% to 200%, (AM-depth from 100% to ∞ %). For values larger than 100%, A_{\max} and A_{\min} always have opposite signs.

If amplitude is defined as maximum departure from a reference value, and modulated signals are represented in two dimensions, the mathematical necessity⁴⁴ of alternately positive and negative amplitude values within the same signal cannot be explained. An alternative would be to use three- rather than two-dimensional signals and understand AM-depth / AF-degree values larger than 100% as resulting in modulated signals that are ‘twisted’ around the x -axis f_{mod} times per second. Signal twists would be accompanied by a rotation of the signals’ coordinate system around the x axis (time) and could explain why the amplitude sign alternates between positive and negative f_{mod} times every second. This explanation will be examined further in Section 6.3.2.

⁴⁴ For AM-depth / AF-degree values $> 100\%$.

5.2.2 Do Amplitude Fluctuations Carry Energy?

Based on their physical characteristics, amplitude fluctuations satisfy all criteria outlined in the proposed definition of sound waves (Section 4.1). Eq. (5.3) describes two sinusoidal vibrations:⁴⁵

$$y_1(t) = A_1 \cos(2\pi f_1 t + \phi_1) \quad \& \quad y_2(t) = A_2 \cos(2\pi f_2 t + \phi_2), \quad \text{Eq. (5.3)}$$

where $y_1(t), y_2(t)$: displacements after time t ; A_1, A_2 : displacement amplitudes;

f_1, f_2 : frequencies with $f_1 < f_2$; ϕ_1, ϕ_2 : initial phases, and $\pi = 3.14$ (let $A_1 = A_2 = A$ and $\phi_1 = \phi_2 = 0$ rads).

Eq. (5.4) describes the displacement of the complex vibration, $y(t)$:

$$y(t) = y_1(t) + y_2(t) = A \cos 2\pi f_1 t + A \cos 2\pi f_2 t =$$

$$2A \cos 2\pi \frac{f_2 - f_1}{2} t \left(\cos 2\pi \frac{f_2 + f_1}{2} t \right) = 2A \cos 2\pi f_{\text{mod}} t \left(\cos 2\pi f_{\text{av}} t \right) \quad \text{Eq. (5.4)}$$

providing that:

$$\frac{f_1 + f_2}{2} = f_{\text{av}} \gg \frac{f_2 - f_1}{2} = f_{\text{mod}} \quad \text{Eq. (5.5)}$$

⁴⁵ For the remaining of this dissertation the ‘real’ portion of a sinusoidal vibration (wave) will be represented by a cosine and the ‘imaginary’ portion by a sine. This is the mathematically correct representation (the solution of the general wave equation is a complex number) and results in the three-dimensional signals proposed in Section 6.3.2. Sines were used up to this point to follow convention and to avoid confusion between mathematical representation (cosine) and terminology (sinusoidal).

Eq. (5.4) indicates that the two-component complex vibration is equivalent to a vibration with frequency f_{av} and amplitude A_{mod} that modulates with frequency f_{mod} ($A_{mod} = 2A \cos 2\pi f_{mod} t$). If the complex vibration is moving in space as a traveling wave (i.e. longitudinal wave in air), Eq. (5.4) describes the displacement around the travelling equilibrium if t is replaced by $t' = t - x_0/c$, where c is the wave velocity and x_0 is the position of the moving particle relative to the equilibrium at time $t_0 = 0$.⁴⁶ In this case, Eq. (5.4) describes modulations with frequency f_{mod} that move along with the wave of frequency f_{av} .

In general, modulations behave like regular waves in all respects (Feynman et al., 1964: 48.5-48.6; Crawford, 1973: 295) and can therefore be considered as energy carriers. They propagate in the same medium as the individual components of a complex wave with the *group velocity* v_{gr} distinguished from the velocity of the components called *phase velocity* v_{ph} . Group and phase velocities are related by the expression:

$$v_{gr} = v_{ph} + \sigma \frac{dv_{ph}}{d\sigma} \quad \text{Eq. (5.6)}$$

⁴⁶ This statement demonstrates that displacement and displacement amplitude can be defined relative to any point of interest rather than to an absolute point of rest. For a proof of this statement using the complex (i.e. including the imaginary part) equation of wave propagation see Feynman et al. (1964: 48.5) or Crawford (1973: 339-341).

where $\sigma = 1/\lambda$ (λ : wavelength) is the *wave index* that describes waves in space just as frequency $f = 1/T$ (T: period) describes waves in time. Group and phase velocities describe two distinct physical variables that, in the case of non-dispersive media (i.e. air) where $dv_{ph}/d\sigma = 0$, have the same value. This does not mean that modulations or fluctuations lose their identity and carry no energy when they propagate in air. It simply means that they propagate with the same velocity as the carrier.

The difference between phase and group velocities can be best understood by observing the propagation of complex waves in a dispersive medium such as water.⁴⁷ Throwing a small stone in a water pond results in circular disturbances spreading away from the point of impact. The waves created by the impact are complex waves with a large number of components. The larger crests ($\lambda_{mod} > 3\text{cm}$) spreading out are modulations (beats, pulses) passing over a fixed point on the surface at a rate equal to f_{mod} . When these modulations reach a leaf resting on the surface they cause it to move, indicating that they carry energy. A closer look reveals that the initially observed disturbances (f_{mod}) are made out of smaller ones ($f \gg f_{mod}$, $\lambda \ll 3\text{cm}$) that move relatively to the modulations. The smaller disturbances start behind the large disturbances, move through them, and die out after the last large disturbance. The smaller disturbances travel with the phase velocity v_{ph} , which, in this example, is

greater than the velocity of the larger disturbances. The velocity of the modulations / beats (larger disturbances) is the group velocity v_{gr} .

It is important to note that the above cannot be observed with such clarity in a non-dispersive medium such as air. This is not because modulations do not behave as waves or carry no energy when they propagate in air, but simply because modulations and carrier travel with the same velocity. Regardless of whether a medium is non-dispersive or dispersive, modulations / beats / envelopes always travel with the group velocity (Elmore & Heald, 1969: 123; Ostrovsky & Potapov, 1999: 8) as does wave energy (Ostrovsky & Potapov, 1999: 63).

Therefore, when two or more sinusoidal vibrations propagate in air as a single complex wave, the resulting modulations / fluctuations represent second order pressure variations (i.e. variations of pressure variations with specified amplitude and frequency values) that propagate with a specified velocity. The fact that amplitude fluctuations of a sound wave in air correspond to second order pressure variations does not indicate lack of energy content. The energy content of modulations in general and amplitude fluctuations in particular can be qualitatively appreciated by reexamining the definition of amplitude. Defining the amplitude of a vibration (or a sound wave) as the maximum displacement of air particles from a point of rest (or the maximum deviation from atmospheric pressure) is problematic since, in the absence of a sound

⁴⁷ In such a medium, modulations separate from the carrier and are therefore easier to observe.

wave, the air particles are still oscillating. Correcting the problem by introducing the requirement for the oscillation to be nonrandom or periodic would exclude almost all sounds since, strictly speaking, no actual sound, pitched or non-pitched, is represented by a truly periodic signal.

All periodicity theories of sound perception rely on the hypothetical ability of the perceptual apparatus to analyze the incoming sound signal in the temporal domain. Such theories assume that the two-dimensional signal in question has a physical reality and/or that listeners are able to perceive small departures from periodicity while ignoring larger ones.⁴⁸ These assumptions are carried forward without suggesting: a) absolute or difference thresholds of periodicity detection or b) why, in spite of our presumed temporal-detection capacity, a far inferior⁴⁹ physical system has developed (i.e. cochlea) with characteristics corresponding to our perceptual abilities better than those of an admittedly more efficient hypothetical temporal-detection mechanism. A principal justification for models based on temporal processing of auditory-nerve firing patterns has been the assumed inability of auditory-nerve discharge rates to represent the dynamic range of the auditory system. Recent discoveries of nerve fibers with broad dynamic ranges are reviving interest in rate-based loudness models

⁴⁸ This refers to claims that the pitch-shift effect (see Section 5.4.2) can be explained in terms of our ability to detect very fine signal variations over time while, at the same time, ignoring much larger such variations a) in noise or b) during the hypothetical development of pitch templates that exhibit remarkable inter- and intra- consistency, despite the variability of experience over time and over different listeners.

⁴⁹ In terms of frequency / intensity / temporal resolution, fatigue limits, masking characteristic, etc.

and location-based pitch models (Sachs, 2001) that do not require the presence of a temporal-fine-structure detection mechanism.

The problem of defining a vibration's absolute point of rest has much in common with the mechanical problem of selecting a stable mode out of all possible modes in a vibrating sphere or the quantum problem of selecting one state as stationary and significant out of all possible electron orbits in the hydrogen atom. Both these problems were solved by employing arbitrary rules to select stable or significant states based on a specific question of interest (Menzel, 1961).

In a similar manner, Federal Standard 1037C (1996) defines amplitude and phase in terms of relative rather than absolute reference points. The reference points are selected based on the source-receiver system of interest and do not have to be fixed but can belong to another varying phenomenon. Based on such a relative definition of amplitude, pressure amplitude fluctuations of sound waves in air are characterized by a pressure amplitude value and therefore carry energy. Evidence of the energy content of amplitude fluctuations has been provided by vibration measurements on machinery gear-boxes (Krishnappa, 1998; Courrech, 1998). While a good gear gives a sinusoidal signal, faulty gears demonstrate amplitude fluctuations caused by misalignment (i.e. Figure 5.4), pitch errors, wear, uneven gear-tooth spaces, or uneven load.

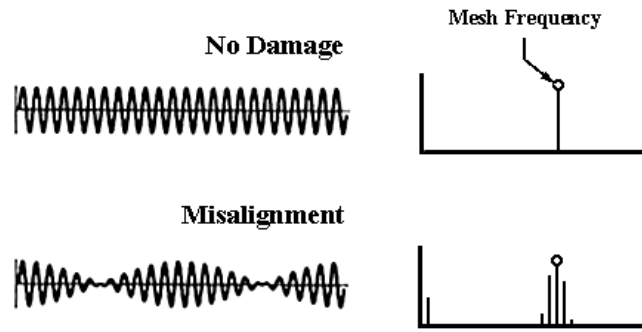


Figure 5.4: Signals and spectra produced by a good and a misaligned gear (after Krishnappa, 1998: 706). Along with the sidebands, the spectrum of the misaligned gear includes a low-frequency component corresponding to the rate of amplitude fluctuation. (*Mesh* frequency is a frequency related to the rotation speed of the gear and to gear-tooth spacing).

Depending on how gradual or sudden those faults are, the fluctuations will have less or more components respectively. These fluctuations can cause serious damage (beyond the breaking of the faulty gear) if the natural frequency of another part of the machinery matches the fluctuation rate (or the rate of any of the amplitude fluctuation's frequency components), indicating that amplitude fluctuations carry energy.

From a theoretical point of view, the contention that amplitude fluctuations do not carry energy but are simply signal variations (Plomp, 1966: 471; Kohda, 1985; etc.) contradicts basic assumptions of physics. Fluctuations in the displacement amplitude of a vibration correspond to velocity fluctuations. Any change in the velocity of some mass requires / reveals the action of a force doing work. How can one expect the velocity of an oscillating particle to change without oscillating energy being drawn / radiated at the rate of velocity changes? Additionally, the phenomenon of diffraction indicates that wave envelopes behave like waves. As mentioned earlier,

in the discussion on Huygens's principle, diffraction can be explained best if successive portions of a wave-front are understood as deriving from the envelope of secondary wavelets. At any given moment, each point on a wave-front is considered a source of secondary wavelets that move away from it. At a later moment, the wave-front is the envelope of the combined secondary wavelets (Elmore & Heald, 1969: 324). Unless we are prepared to accept that diffracted sound waves carry no energy, amplitude modulations / fluctuations must be considered energy carriers.

Further theoretical arguments and empirical evidence demonstrating that modulations in general and amplitude fluctuations in particular carry energy are presented in the discussion on the difference tone (Section 5.4). Section 6.3 examines in more detail the issue of the energy content of amplitude fluctuations.

5.3 Roughness: Models Quantifying Roughness and their Drawbacks

5.3.1 Introduction

The present study has approached the sensation of roughness as one of the perceptual manifestations of amplitude fluctuation, associated with a certain range of amplitude fluctuation rates. Examination of performance practices and musical instrument construction from around the world (Chapter 2) demonstrated that sound variations involving this sensation are found in most musical traditions. In the Western musical tradition, the sensation of roughness⁵⁰ has often been linked to the concepts of consonance and dissonance, whether those have been understood as aesthetically loaded (Rameau, Romieu, in Carlton, 1990; Kameoka & Kuriyagawa, 1969a; Terhardt, 1974a&b, 1984) or not (Helmholtz, 1885; Hindemith, 1945; von Békésy, 1960; Plomp & Levelt, 1965). Studies addressing this sensation have occasionally been too keen to find a definite and universally acceptable justification of the 'natural inevitability' and 'aesthetic superiority' of Western music theory (i.e. Stumpf, 1890, in von Békésy, 1960: 348; Vogel, 1993; etc.). This has prevented them

⁵⁰ Similarly to the sensation of beats.

from seriously examining the physical and physiological correlates of the roughness sensation. On the contrary, Helmholtz (1821-1894), the first researcher to examine roughness theoretically and experimentally as an important attribute of auditory sensation, concluded:

Whether one combination [of tones] is rougher or smoother than another depends solely on the anatomical structure of the ear, and has nothing to do with psychological motives. But what degree of roughness a hearer is inclined to ... as a means of musical expression depends on taste and habit; hence the boundary between consonances and dissonances has frequently changed ... and will still further change... (Helmholtz, 1885: 234-235).⁵¹

After Helmholtz's work, roughness got little attention in psychoacoustics⁵² until the 1960s when studies by von Békésy, Terhardt, Plomp, and others acknowledged roughness as one of the main auditory attributes along with pitch, loudness, and timbre.⁵³ Since then, further studies have demonstrated that the sensation of roughness plays an important role in several aspects of sound evaluation. Within the Western tradition, the consonance or 'absence of annoyance' in non-musical sounds has also been shown to depend on roughness (Cardozo & van Lieshout, 1981;

⁵¹ Similarly, Ortmann (1922) concluded that the most important factor in the enjoyment of music is non-auditory, with roughness being much less important than the listener's associations when it comes to music evaluation and appreciation.

⁵² With the exemption of von Békésy's work in the mid 1930s.

⁵³ Lichte (1941) was one of the first to consider roughness one of the primary attributes of complex tones. The sensation of roughness is not necessarily associated with combinations of notes. It can arise from the interference among the components of a single complex tone or from performance practices in monophonic music (i.e. fast vibrato). The only condition is that the resulting complex signal exhibits amplitude fluctuations within a specified range of fluctuation rates (~ 20–150 fluctuations per second).

Terhardt & Stoll, 1981; Vos & Smoorenburg, 1985; Hashimoto & Hatano, 1994).⁵⁴

Along similar lines, Pressnitzer et al. (2000) confirmed that, as the perceptual salience of other attributes (such as high pitch or tonal character) is reduced, the correlation between nontonal tension and roughness increases.

The present study treats the sensation of roughness as a perceptual manifestation of the energy content of amplitude fluctuations, which can be manipulated by controlling the fluctuation rate and degree, providing means of sonic variation and musical expression.

⁵⁴ Applications of roughness evaluation are also found in medicine (i.e.: Imaizumi, 1986; Zeh et al., 1987; etc.).

5.3.2 Models Quantifying Roughness and their Drawbacks

The two principal studies that have systematically examined the sensation of roughness (von Békésy, 1960: 344-354, Terhardt, 1974a) have, to a large extent, been ignored by the models quantifying auditory roughness / smoothness of complex spectra. Numerous such models have been proposed over the last ~100 years (i.e. Helmholtz, 1885; Plomp & Levelt, 1965; Kameoka & Kuriyagawa, 1969a&b; Hutchinson & Knopoff, 1978, Sethares, 1998) and have been employed in later studies (Bigand et al., 1996; Vos, 1986; Dikken, 1999, etc.) demonstrating a relatively low degree of agreement between predicted and experimental data.

Dikken, for example, found no correlation between sensory consonance⁵⁵ (smoothness), as predicted by the Hutchinson & Knopoff model, and the completeness / stability ratings of the final bars of selected musical pieces. She concluded that sensory consonance / dissonance is not a good measure of musical stability / tension, or completeness / incompleteness, interpreting her conclusion as supporting the need for an alternative model of consonance / dissonance. Her study is a good example of an attempt to examine questions that go beyond the scope of the model employed. It

⁵⁵ The term *sensory consonance* was introduced by Plomp & Levelt (1965) to distinguish consonance understood simply as absence of roughness from the aesthetically loaded concept with the same name.

essentially demonstrates that the degree of smoothness of a vertical sonority (i.e. chord) is not a good measure of its degree of stability or completeness. This result could have been anticipated since the stability of any given event is temporal in character and therefore highly related to the events that precede or follow it, while roughness models calculate the roughness of isolated vertical sonorities. The surprising fact is not Dibben's results but the implied expectation that a measure of smoothness could correlate with multidimensional, highly temporal, and context-dependent notions such as stability or completeness.⁵⁶ It appears that the concept of consonance, like that of timbre, has been a wastebasket⁵⁷ of aesthetic and evaluative judgements in music, as well as the source of justification arguments regarding general stylistic trends or specific compositional decisions.

Nonetheless, there are many reasons (other than those suggested by Dibben) for the revision of the existing models quantifying roughness. Some have already been pointed out by other researchers and several will be addressed by the present study. Helmholtz's model (1885) was abandoned after perceptual experiments (i.e. Plomp & Levelt, 1965) revealed a nonlinear relationship between roughness and

⁵⁶ Bigand et al. (1996) were more successful in their application of the Hutchinson & Knopoff roughness model because they combined it with the concept of *horizontal motion*, tracking roughness changes with time rather than computing isolated roughness values.

⁵⁷ This refers to Bergman's (1990: 93) assessment that timbre seems to be the wastebasket of all sound characteristics that cannot be placed under the labels of pitch or loudness.

register.⁵⁸ Vos (1986) pointed out a number of inconsistencies in the Plomp & Levelt (1965) and the Kameoka & Kuriyagawa (1969a&b) models,⁵⁹ with regards to the critical bandwidth model derived from loudness summation experiments (Zwicker et al., 1957). In his study, Vos suggested adjustments that would bring the predictions of all three models to a closer agreement. Hutchinson & Knopoff's model (1978) has been criticized (Bigand et al. 1996) for its somewhat crude representation of the observed nonlinear relationship between register and the amplitude fluctuation rate corresponding to maximum roughness.

A recent model by Sethares (1998) has the advantage of being based on a large number of smoothness / roughness experimental ratings of sine-pairs, fitting a function that accounts for the previously mentioned nonlinear relationships. A drawback this model shares with all earlier ones is that it misrepresents the contribution of the amplitudes of the interfering sines (and therefore of the AF-degree of the resulting complex signal)⁶⁰ to the degree of roughness. Usually, the roughness function for a

⁵⁸ In his model, Helmholtz (1885: 192) assumed maximum roughness of a pair of sines at a fixed frequency difference (33Hz) for all registers.

⁵⁹ The Plomp & Levelt (1965) model underestimates roughness because of its bias *against* a power function for roughness, while the Kameoka & Kuriyagawa (1969a&b) model overestimates roughness because of its bias *for* a power function for roughness. That some sort of power function (although not exactly the one relating amplitude to loudness) is called for is supported by observations indicating a relationship between the mechanisms associated with the sensations of roughness and loudness (Von Békésy 1960: 344-350).

⁶⁰ If two sines with different frequencies f_1, f_2 ($|f_1 + f_2|/2 \gg |f_1 - f_2|/2$) and amplitudes A_1, A_2 ($A_1 \geq A_2$) are added together, the amplitude of the resulting signal will fluctuate between a maximum

sine-pair is multiplied by the product of the two amplitudes ($A_1 A_2$), ensuring minimum roughness if either of the amplitudes approaches zero. At the same time, however, it severely overestimates the increase in roughness with increasing amplitudes and, most importantly, it fails to capture the relationship between the amplitude difference of two sines close in frequency and the salience of the resulting beats or roughness.

As mentioned in Section 3.4, Terhardt (1974a) examined experimentally the influence of modulation depth m and sound pressure level SPL on the roughness of amplitude modulated tones, as well as the relationship between the roughness of amplitude modulated tones (modulation frequency f_{mod} , modulation depth $m = 1$) and the roughness of sine-pairs with amplitude fluctuations of the same rate ($|f_1 - f_2| = f_{mod}$) and degree ($A_1 = A_2$). Terhardt examined the relationship between degree of amplitude fluctuation and roughness through a manipulation of amplitude modulation depth, assuming $AF\text{-degree} = AM\text{-depth}$. Chapter 3 showed that this assumption is incorrect. Therefore, the functions introduced by Terhardt to describe the above relationships were adjusted accordingly. The adjusted equations from Section 3.4 indicate that:

($A_{max} = A_1 + A_2$) and a minimum ($A_{min} = A_1 - A_2$) value. As indicated in Section 3.2, the degree of amplitude fluctuation (AF_{degree}) is defined as the difference between the maximum and minimum amplitude values relative to the maximum amplitude value. So, $AF_{degree} = (A_{max} - A_{min}) / A_{max} = 2A_2 / (A_1 + A_2)$.

- a) the contribution of sound pressure level to roughness is negligible and
- b) the contribution of the degree of amplitude fluctuation to roughness is larger than assumed by all existing models calculating roughness⁶¹ and even larger than indicated by Terhardt (1974a).

Based on the necessary adjustments, an appropriate variable can be added to Sethares' model to account for the contribution of amplitude to roughness, completing a new model for the roughness estimation of pairs of sines (Section 6.4). The roughness of complex spectra with more than two sine components will be estimated by adding together the roughness of the individual sine-pairs. Although it has been argued that the total roughness can be less than the sum of the roughness of each sine-pair⁶² (von Békésy, 1960: 350-351), initial experiments indicated otherwise confirming results from previous studies (Terhardt, 1974a; Lin & Hartmann, 1995).⁶³

⁶¹ Hutchinson & Knopoff (1978) assume a linear relationship between amplitude fluctuation degree and roughness, while all other models completely ignore the degree of amplitude fluctuation from their calculations.

⁶² Depending on the relative phase of the respective amplitude fluctuations.

Correct estimation of the roughness degree of a pair of sines or of any arbitrary spectrum is necessary before some claimed link between roughness and an acoustic, perceptual, or musical variable can be tested. Developing a detailed model that estimates the roughness of complex spectra (Section 6.4) permitted examination of the relationship between the sensation of roughness and the harmonic-interval hierarchy suggested by Western music theory. It is argued that this hierarchy represents Western musical tradition's subtle attempt to exploit one of the perceptual attributes of amplitude fluctuation within an aesthetic that, for the most part, considers roughness as dissonant.

⁶³ More specifically, Lin & Hartmann (1995) concluded that the total roughness is summed over all auditory filters.

5.4 The Difference Tone

5.4.1 History, Physical Evidence, Explanations

Although the phenomena of beating and of the difference tone (or, more generally, of combination tones) were first observed separately, their theoretical examination has had a parallel development. The interest of musicians, as well as of acousticians, on beats and combination tones was initially associated to the long-lasting involvement of theoretical music with numerical theory and small integer ratios. Although this involvement continues for some (i.e. Vogel, 1993), since Helmholtz's work both phenomena have been examined mainly with regards to their physical and perceptual properties. Along with the beating and roughness sensations, the difference and other combination tones have since been linked to music theory and practice through perception rather than mathematics.

The term *difference tone* describes a tone with pitch corresponding to the frequency $|f_1 - f_2|$, heard when two tones with fundamental frequencies f_1 and f_2 are played together.⁶⁴ First reported in the middle of the 18th century, the difference tone

⁶⁴ The frequency that matches in pitch the difference tone is therefore equal to the amplitude fluctuation rate of the resulting two-component tone complex.

was announced independently by a) Sorge (1744), a German organist and music theorist, b) Romieu (1751), a French organist, known mainly for an extensive study on tuning and temperament, and c) Tartini (1754) an Italian violinist, composer, and theorist.⁶⁵ Tartini's announcement received most of the attention since it was the only one presented in the form of a formal treatise, introducing the difference tone as the foundation of a new harmonic theory. None of the tone's discoverers, however, offered a physical theory to explain its origin. Tartini, in *Trattato di Musica*, offered metaphysical explanations, as did Sorge who spoke of the "resolution of multiplicity [of the upper components of a harmonic series] into ... a fundamental unity" (in Carlton, 1990: 43-44).

During the first half of the 19th century, Lagrange, Chladni, and Young studied the difference tone concluding it was a form of beating. It was argued that, as the frequency separation of the original tones increases, the rate of amplitude fluctuation resulting from their interference moves into the audible range ($> \approx 50\text{Hz}$) and the sensation changes gradually from beating to roughness and finally to a new tone, the difference tone. Since then, there has been an ongoing controversy regarding whether

⁶⁵ The order of the discovery presented here follows Carlton (1990). Other historians argue for the reverse order (i.e. Jones, 1935, in Plomp, 1965: 1111). None of the discoverers of the difference tone referred to it by that name. Sorge gave it no particular name, Romieu called it *grave harmonic* and Tartini called it the *third sound*. Neither term was adopted. Due to the publicity given to Tartini's theory (for various reasons) over an earlier and better-formulated similar theory by Rameau (early 1700s), the discovered tone was termed *Tartini tone*, a term that still survives. In 1805, Vieth coined the term *combination tone* and finally Helmholtz, in his study "On Combination Tones" (1856), introduced the term *difference tone* to distinguish it from another combination tone he discovered, the *sum tone*.

the difference tone in particular and combination tones in general exist objectively as pressure variations in the air, whether they are the products of nonlinear response of the ear, or products of some higher level (neural) processing.⁶⁶ Von Békésy (1960) has provided experimental evidence of the nonlinear production of the difference tone inside the cochlea. These evidence, although often interpreted as such, do not question difference tone's objective nature. They simply show that the difference tone may be reinforced due to the nonlinear characteristics of the inner ear. Helmholtz was aware of the possibility of the difference tone existing outside the human ear and was able to detect it by setting up sympathetic vibrations in thin membranes (Helmholtz, 1885: 531). Further experimental verifications of energy at the difference frequency are outlined below:

- Rücker & Edser (1895, in Beyer, 1999: 149-150) used a siren to produce two tones with a frequency difference of 64Hz, activating (via a Koenig resonator) a tuning fork tuned at the same frequency, even for small amplitudes of the interfering tones.
- Jeans (1937) struck C3 and G3 simultaneously on the piano and observed that "... *we not only hear C2* [tone corresponding to the frequency difference between C3 and G3], *but (...) also find that the C2 strings are feebly vibrating, as can be*

⁶⁶ For Rayleigh there was no question: "a difference tone is not a mere sensation but involves a *vibration* of definite amplitude and phase." (Rayleigh, 1896: 462).

verified by laying small chips of wood across them ...” (Jeans, 1937: 238). In his discussion on double-pipe whistles Jeans remarks: *"The pipes may even be so short that their individual tones are above the range of audition. Then neither pipe can be heard when sounded separately, but when the two pipes are sounded together, the difference tone is heard quite clearly."* (Jeans, 1937: 242).⁶⁷

- Dolphin et al. (1994) recorded auditory-evoked potentials from the scalp, showing the existence of a physical component at the difference frequency.
- Howard (1989) found that hearing aids furnishing fundamental frequency information through waveform peak-picking, produce better scores in profoundly hearing impaired subjects than amplification (as compared to a relevant study by Faulkner et al., 1992). His study indicates that beats, resulting from the interference between successive components of complex tones, carry important information regarding pitch and may constitute actual spectral components.
- Rayleigh (1896), after discussing the experimental observations of Helmholtz and Rücker & Edser, stated: *"My own observations have been made upon the harmonium, and leave me at a loss to understand how two opinions are possible [regarding the objective nature of the difference tone, even for small amplitudes of*

⁶⁷ König & Mayer reported the same observation in 1894 (in Rayleigh, 1896: 462).

the two interfering tones]. *The resonator is held with its mouth as near as may be to the reeds which sound the generating notes, and is put in and out of tune to the difference tone by slight movements of the finger. When the tuning is good the difference-tone swells out with considerable strength, but a slight mistuning (...) reduces it almost to silence.*" (Rayleigh, 1896: 457). Later on he argued that for all waves propagating in air "*there is an essential want of symmetry between condensation and rarefaction, and the formation in some degree of (...) combination tones is a mathematical necessity*" that depended on certain unspecified, for Rayleigh, conditions (Rayleigh, 1896: 464).

The 'essential asymmetry' Rayleigh is referring to was first noted formally by Poisson (1808, in Beyer, 1999: 40) and was expressed mathematically by Riemann (1860, in Beyer, 1999: 148-149). Poisson argued that the local particle velocity (i.e. displacement velocity) always affects the velocity of propagation of a longitudinal wave. The general form of a plane-wave particle velocity should therefore be $v = F[x \pm (c_0 + v)t]$ rather than $v = [F(x \pm c_0 t)]$, where F is some function of the momentary displacement x and of the velocity of propagation c_0 . In other words, for the effective velocity of propagation c_{ef} we always get $c_{ef} > c_0$ in condensations and $c_{ef} < c_0$ in rarefactions. It is this asymmetry that Helmholtz modeled by including a quadratic term in the equation of wave motion, revealing the mathematical necessity, under certain conditions, of the difference tone (1856, in Beyer, 1984: 229).

Westervelt (1957, in Beyer, 1999: 317) identified collinearity rather than large amplitude of the interfering sources as a principal such condition. Crighton (1998: 190) states that, regardless of the amplitude of the constituent sines, the propagation of complex signals exhibits linear behavior only locally (i.e. at the scale of the frequency of the carrier), while it exhibits nonlinear behavior over long / slow scales (i.e. at the scale of the amplitude fluctuation rate).⁶⁸ Hamilton (1998: 221-224) provides a summary of Westervelt's results, giving a number of formulas that correspond to different amplitude distribution profiles (i.e. radiation patterns) of a source. These formulas calculate the linear approximation of the pressure amplitude radiated at the difference frequency⁶⁹ and provide the theoretical justification to the results obtained a century earlier by Rücker & Edser. They indicate that, for Gaussian sources the nonlinearly generated difference tone forms a narrower beam than if it were radiated directly by the source of the primary beams.⁷⁰

Several issues regarding combination tones have been addressed by physical acoustics in the study of *scattering of sound by sound*. Westervelt based his theory of

⁶⁸ Large amplitudes of the interfering sources, listed in the psychoacoustics literature as a necessary condition for the nonlinear creation of difference tones, is only necessary to the nonlinear creation of difference tones inside the ear.

⁶⁹ For sound beams with two components originating from a single source.

⁷⁰ This may explain why in Rayleigh's experiments the difference tone fell off with distance much sooner than a generated tone of equal amplitude and frequency (1896: 460-461).

mutual scattering (developed through studies outlined in Beyer, 1984) on earlier theoretical work by Lamb (1931) and on experiments by Thuras et al. (1935) (both in Beyer, 1984) examining the case of two collinear sound waves in a pipe. This theory treats the difference tone produced by the interference of two collinear sound beams as a pressure wave and names it *parametric array*. The new name was introduced to highlight the similarity between the difference frequency and the *endfire array* (effect related to underwater signaling techniques) in terms of a) narrow directivity at low frequencies and b) low absorption.⁷¹

While techniques that exploit the difference frequency have found wide applications in physical and underwater acoustics, musical and psychological acoustics insist on questioning the objective nature of the difference tone, failing to appreciate its perceptual and musical significance. One of the reasons for this insistence is the phenomenon of the *pitch-shift effect* addressed in the following Section.

⁷¹ Westervelt, 1963, in Beyer, 1999: 317.

5.4.2 The Pitch-Shift Effect

In spite of the evidence demonstrating the objective nature of the difference tone, the link between the difference tone and amplitude fluctuation has been contested for two reasons:

- a) Observations from dichotic experiments suggest that sound interference products may arise from neural wave interaction, questioning the significance of physical interference to perception. The issue of dichotic interference is addressed in Section 5.5.
- b) The perceptual phenomenon known as the *pitch-shift effect* indicates that the rate of amplitude fluctuation does not always correspond to the pitch of the difference tone, questioning the link between the two phenomena.

The term *pitch-shift effect* is used in the present study to refer collectively to two phenomena that, until now, have been considered as discrete: the *first* and the *second pitch-shift effects*.

The *first pitch-shift effect* describes changes in the pitch of a complex harmonic signal, resulting from the uniform frequency shift Δf of all its components. De Boer (1956a&b) conducted one of the first systematic studies examining this phenomenon, with stimuli consisting of 5 or 7 components spaced at $f_0 = 200\text{Hz}$ apart

and shifted by an equal amount of Δf Hz. The perceived pitch of each stimulus followed a saw-tooth-like function around the fixed frequency spacing f_0 , across the range of center frequencies (Figure 5.5).

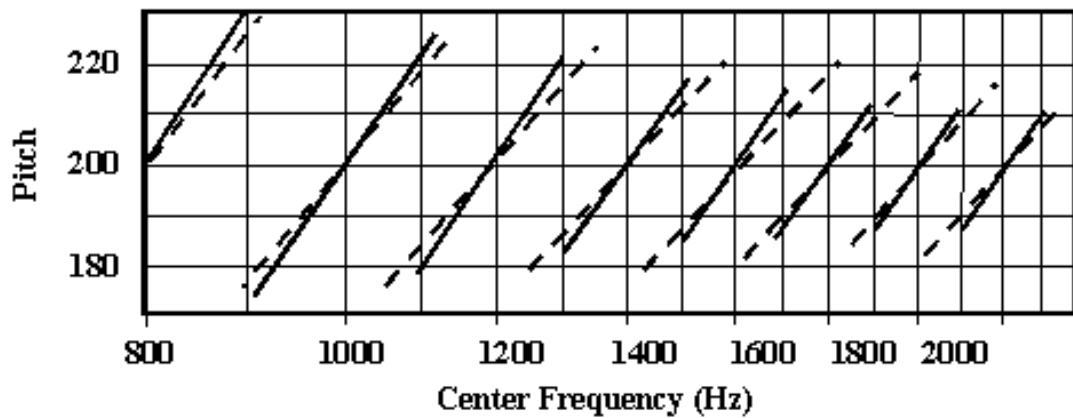


Figure 5.5: Illustration of the *first pitch-shift effect*. Dependence of the pitch (abscissa) of a complex stimulus with 5 equally spaced components (frequency spacing: $f_0 = 200\text{Hz}$) on the frequency (ordinate) of its center component (after de Boer, 1956b).

The continuous lines (Figure 5.5) describe the observed relationship between pitch and center frequency while the broken lines describe the same relationship as predicted by the mathematical model proposed by de Boer to describe the phenomenon.

Since de Boer, various relevant studies have been conducted⁷² largely confirming his results.⁷³ In order to account for the absolute value of the observed

⁷² i.e.: Schouten et al. (1962), Schroeder (1966), Smoorenburg (1970), etc.

⁷³ For a slightly different interpretation of these results see van den Brink (1970).

pitch shift being consistently larger than the one predicted by de Boer's model, all studies have introduced some adjustment to this model, supported by various theoretical arguments. Smoorenburg (1970) presented the most convincing argument, demonstrating that the effective center frequency of a stimulus is reduced by the presence of combination tones,⁷⁴ resulting in a larger absolute value for the predicted pitch shift, a value that agrees with experiment.

The *second pitch-shift effect* describes a slight drop / rise in pitch when the frequency spacing between the components of a harmonic signal is slightly increased / decreased respectively, while keeping the center frequency fixed (Figure 5.6).

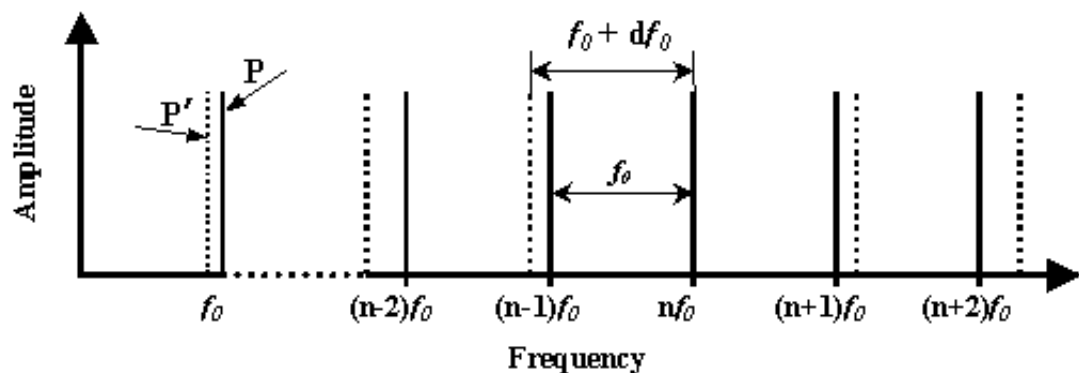


Figure 5.6: Illustration of the *second pitch-shift effect*. Increasing the spacing of the components around a fixed center component n , results in a drop in pitch $P' < P$, although the difference frequency (frequency spacing) has increased from f_0 (continuous lines) to $f_0 + df_0$ (broken lines).

⁷⁴ Smoorenburg referred to combination tones of the type: $f_i - k(f_2 - f_1)$. [f_1, f_2 : Successive components, $f_1 < f_2$, $f_2/f_1 < 2$, and k : small integer]. His results indicated that combination tones contribute to the pitch of complex tones as if those frequencies were present in the stimulus.

Vassilakis (1998) showed that the *second pitch-shift effect* is an unnecessary concept since the phenomenon it describes is already predicted by the *first pitch-shift effect*. A single model was therefore introduced, outlining what the present study will be referring to simply as the *pitch-shift effect*.

Essentially, the pitch-shift effect represents a discrepancy between the rate of amplitude fluctuation and the frequency corresponding to the pitch of inharmonic complex signals (i.e. complex signals whose components have frequencies that are not integer multiples of a fundamental) with equally spaced frequency components. This phenomenon must not be interpreted as questioning the energy content of amplitude fluctuation. Pitch perception of complex tones appears, according to a number of additional observations (Deutsch, 1999a&b), to be based on interpretation rather than straight translation of spectral distributions. It involves memory, previous learning, and, most importantly, contextual information such as the presence of other complex tones, both simultaneous and in succession.⁷⁵ Hence, the difference between the frequency corresponding to the pitch of some complex tones and the rate of amplitude fluctuation of the related signals questions the energy content of the amplitude

⁷⁵ For example, a specific frequency component may be interpreted, after it has already subsided, as the fundamental of a complex tone that follows it. The conditions are that the second tone is made out of a wide band of noise and a single component, possible harmonic of the previously-heard frequency component, and has a low signal to noise ratio (Houtgast, 1976).

fluctuations no more than it questions the energy content of any other spectral component of the tone.

Theories of pitch perception range from physiological and neural (i.e. place and periodicity theories), to autocorrelation, pattern recognition, and template theories (i.e. optimal processor and virtual pitch theories), all of which attempt to fit their models to observation, while failing to adequately account for one or more of the following:

- a) The pitch-shift effect.
- b) Binaural diplacusis, describing frequency-pitch differences for a person's two ears.
- c) Pitch shifts of pure, as well as complex tones (with or without energy at the fundamental) resulting from the introduction of white noise, band-filtered noise, or high- / low-pass-filtered noise.
- d) Non-periodic complex tones with prominent pitch (i.e. narrow bands of noise).
- e) Complex signals with the same periodicity and different pitches, or the same pitch and different periodicities.
- f) Perception of the missing fundamental (or residue pitch)⁷⁶ in cases where the claim of template theories regarding "substantial learning during the listener's childhood" is not justified (i.e. in animal experiments).

⁷⁶ This refers to the observation that the pitch of a harmonic complex tone matches in frequency the fundamental component, even if this component is not actually present in the tone's spectrum. The term *residue pitch* was introduced by Schouten et al. (1962) as part of a theory of pitch based on the periodicity properties of the non-resolved or "residue" components of a complex tone. The theory was formulated in efforts to explain the pitch of the missing fundamental.

g) Auditory filter with a much coarser frequency resolution than supported by an autocorrelation model, and a highly complex physiological construction (inner ear), which would be unnecessary to a hearing mechanism that could perform autocorrelation on incoming sound signals. The eardrum or any uniform, stretched membrane would suffice.

Spectral composition is just one of the cues informing pitch perception. The pitch-shift phenomenon confirms the difference tone as an important component of this cue rather than questions it. As already mentioned, the mathematical model describing the pitch-shift effect can fit to observation only if the difference tone is considered as belonging to the spectra of the stimuli in question (Smoorenburg, 1970). At the same time, the pitch associated with such spectra cannot be linked to any single spectral component but only to the entire spectral distribution. For these reasons, the pitch-shift effect does not question the reality of the difference tone and of amplitude fluctuations any more than it questions the physical reality of all other components of a complex tone, objective or subjective.

Resolving the issue regarding the physical nature of the difference tone and its relationship to signal amplitude fluctuations would provide partial justification to studies that have linked the difference tone to musical choices within the Western musical tradition (i.e. Helmholtz, 1885; Hindemith, 1945; Campbell & Greated, 1987; etc.).

5.5 Dichotic Interference

As already mentioned, the issue of the energy content of amplitude fluctuations is considered insignificant to perception or is altogether dismissed mainly because of reports from dichotic experiments. In such experiments two sine tones with slightly different frequencies are presented one in each ear through headphones, without being allowed to physically interfere. The fact that subjects continue to report beating sensations (Plomp, 1966, 1976; Sethares, 1998; etc.) has been interpreted as evidence of interference products arising from neural wave interaction, questioning the importance of linking physical interference to issues of timbre or pitch. The reported beating sensation has been given quite diverging descriptions, ranging from periodic fluctuations in loudness, timbre, and pitch (Plomp, 1966: 464), to periodic rotations of the sound image (Hornbostel, 1923, in von Békésy, 1960: 349), while the fluctuations are always weaker than in the case of physical interference. From very early on, however, it was noted that this beating-like sensation may be quite different from the beating that arises from physical wave interaction, since dichotic experiments never gave rise to combination tones (Rayleigh, 1896: 440).⁷⁷

⁷⁷ Arnold & Burkard (2000) have reported a dichotic difference tone response recorded from the inferior colliculus in the chinchilla that, according to the authors, is not attributable to cross hearing (physical transfer of a stimulus across ears). Further study is needed.

The present study questions the claim of sound interference products arising at a neural level. If the beating sensation reported in dichotic experiments was a product of interference, Unisons presented dichotically, first in and then out of phase would exhibit a decrease in loudness. However, pilot dichotic experiments using 200Hz sine tones, first in and then 180^0 out of phase, revealed no noticeable change in loudness but an increase in the perceived space occupied by the out-of-phase stimulus. At the same time, if sound interference products could arise beyond the cochlea, then all dichotically presented Unisons would give rise to the sensation of beating since, for the same frequency, there is always a pitch discrepancy between a person's two ears (von Békésy, 1963a; van den Brink, 1970). This physiologically-based phenomenon (binaural diplacusis) results in normal discrepancies of up to ~ 1 whole tone, while, in pathological cases, the pitch difference between the two ears may exceed an Octave. In pilot experiments with subjects reporting binaural discrepancies of less than a semitone, although at a neural level the tones were always slightly detuned, no loudness fluctuation was reported, challenging again the claim of interference products arising beyond the cochlea.

The present study argues that beating-like sensations reported in earlier dichotic experiments can be explained in terms of a) stimulus transfer across ears or *cross hearing*, and b) sound localization cues based on phase differences between the sound arriving in the left and right ears.

Numerous studies (Lane, 1925; Zwislocki, 1953; Nixon & von Gierke, 1959; von Békésy, 1960: 127-203; Nixon & Berger, 1998) have examined cross hearing, both in the free field and through headphones, indicating that the major cross hearing parameter is bone conduction. Tests have included vibration measurements on artificial and actual heads, as well as experiments determining the threshold of hearing for vibrations applied directly on the head or for stimuli presented to patients that were completely deaf in one ear.⁷⁸ Additionally, dichotic masking experiments have shown that for a loud tone of one frequency to mask a faint tone of another frequency, the loud tone must be about 50dB stronger when presented to the ear opposite the faint tone than when presented to the same ear (Wegel & Lane, 1924, in Lane, 1925: 404). The sound level reaching the cochlea via bone conduction has therefore been determined to be approximately 40dB–55dB below the level reaching the cochlea through the ear canal, and is higher for lower frequencies. This may be one of the reasons why beating in dichotic experiments was reported mainly for low-frequency stimuli. Additionally, bone-conduction efficiency may have been increased further in these experiments because of the occlusion effect.⁷⁹ Von Békésy (1960: 176) showed that maximal minimization of cross hearing occurs for stimuli presented through

⁷⁸ A stimulus was presented to the deaf ear and the minimum sound level of the stimulus was determined for which a sensation would register on the healthy ear.

⁷⁹ The term *occlusion effect* refers to a marked increase in the efficiency of bone conduction for frequencies below 2000Hz when the ear canals are occluded with earmuffs or headphone cups. (Nixon & Berger, 1998: 21.5).

headphones that are completely inserted into the ear,⁸⁰ or have relatively large internal volume (i.e. with cushions around the ears). Pilot dichotic experiments conducted for the present study indicated no loudness fluctuation when low-level stimuli and appropriate headphones were used.

However, a beating-like sensation did persist. This sensation was described as a timbral fluctuation that, when the frequency difference between the sines presented in each ear was reduced down to 1Hz, was identified as a stimulus rotation between the subject's two ears. Sound localization studies in the free field have indicated that, for frequencies below 500Hz, the most important directional cues on the horizontal plane come from signal phase-differences between the two ears (Blauert, 1996; Hartmann, 1999: 24).⁸¹ These observations suggest that, in dichotic explorations of beats, the constantly shifting phase between the two stimuli may result in a constantly shifting localization of the complex stimulus. For low frequency differences, this shift may be perceived as a rotation of the sound within the subject's head and, for higher frequency differences, as a timbral fluctuation. In other words, the beating-like sensation reported in previous dichotic experiments may not be an indication of sound

⁸⁰ The sound was fed to the ear through a plastic tube, leading through a perforated earplug. To prevent tube resonance, a thread was placed inside the tube.

⁸¹ Rayleigh (1896: 441-443) was one of the first researchers to theoretically link the perception of direction for low frequency sounds to phase. Since then, research has confirmed that humans and most other animals localize a sound source by noting differences between the two ears in arrival time (equivalent to phase differences) and in intensity (Narins, 2001). Recently, Mason et al. (2001) showed that this is also the case for the small parasitoid fly, the *Ormia ochracea*.

interference products arising at a neural level but a manifestation of a sound localization mechanism based on phase difference cues.⁸² Since this mechanism works poorly for frequencies beyond 1500Hz (Hartmann, 1999: 25) it is possible to test the above claim using low-level, high frequency stimuli. In a pilot experiment, a 2000Hz sine tone was presented to the left ear and sine tones between 2001Hz and 2010Hz were presented to the right (at 1Hz intervals). No beating-like sensation was reported, supporting the claim against the hypothesis of neural wave interaction giving rise to interference products.⁸³ Additionally, the earlier mentioned increase in the perceived space of dichotic 180° out of phase Unisons agrees with sound localization experiments that test the efficiency of phase difference cues. In such experiments, when the signals in the two ears are 180° out of phase, there is maximum localization ambiguity and the sound appears to be at the center, but fuller (Hartmann, 1999: 25).

The general issue of dichotically-presented sine tones giving rise to sound interference products at some level beyond the cochlea is important and will be examined experimentally in the following chapter.

⁸² Fenney (1997) reached a similar conclusion regarding dichotic beats of mistuned consonances (i.e. beating-like sensation arising when two sines with frequencies slightly removed from a small integer ratio relationship are presented dichotically). Detection of dichotic beats of mistuned consonances (even when binaural interaction within a critical band was inhibited through masking) was interpreted as evidence of dichotic beats being a between-channel phase effect.

⁸³ Other than in cross hearing cases, dichotic beats have been reported for frequencies above 1500Hz only for stimuli that were amplitude modulated at rates below 640Hz and different for each ear (Bernstein & Trahiotis, 1996). In such cases, the beating-like sensation is caused by the constantly shifting phase relationship between signal envelopes rather than signals and can be explained similarly to the dichotic beats reported for frequencies below 1500Hz.

Chapter 6 Theoretical Models - Hypotheses - Experiments

6.1 Introduction - Apparatus

Chapters 3 to 5 outlined the need for the present study, identified the study questions, and pointed towards possible answers. Chapter 6 restates the study questions in the form of three related but distinct groups of hypotheses designed to test the proposed answers. The first group of hypotheses challenges the claim of sound interference products arising from neural wave interaction through a series of dichotic experiments. The second group models the energy content of amplitude fluctuations resulting from the addition of two sines, introduces an alternative sound-signal representation that can account for the energy carried by amplitude fluctuations, and tests the proposed model experimentally. The third and final group of hypotheses examines the sensation of roughness, the most widely exploited perceptual attribute of amplitude fluctuation in music. A new model estimating the auditory roughness of complex spectra is introduced. The model is applied to a hypothesis that links the dissonance hierarchy of harmonic intervals in the Western musical tradition to variations in roughness degrees.

In experimental design, categorization and matching tasks have been preferred over identification ones. Additionally, whenever semantic differential ratings were required, they were performed on relative rather than on absolute scales. For example, instead of subjects deciding how loud or soft a stimulus was, they decided how much louder or softer a stimulus was relative to a comparison stimulus.

The stimuli for all perceptual experiments were synthesized digitally on a *Pentium PC* using a program written in *Visual Basic 6.0* (Microsoft Corp., 1998) and shared the following attributes: *Sampling rate*: 44,100 samples/sec; *Bit rate*: 16 bits; *Length*: 1 sec; *Attack*: 22ms; *Decay*: 32ms. Stimuli were fed to a *Rotel RA-931 Stereo Integrated Amplifier* and were presented to the subjects through a pair of *AKG K-240* (600ohms) headphones. The stimuli levels were measured and maintained using a *Fluke 8050A Digital Multimeter*. The average level was 0.28 Volts \pm 0.01Volts (68dBm \pm 1dBm).⁸⁴

All perceptual experiments were designed and conducted on a *Pentium PC* using *MEDS* (Music Experiment Development System) by Kendall (2001). Experiment implementation, including instructions to the subjects, presentation of the stimuli and of the experiment questions, and recording of the responses was

⁸⁴ I.e. 68dB above 1 milliwatt. Reference reading: Headphone jack output for no sound input.

automated. Experimenter interaction with the subjects was limited to offering initial instructions and administering training sessions.

All subjects were graduate students at UCLA's music, musicology and ethnomusicology departments and reported normal hearing. In all experiments, subjects were able to familiarize themselves with the relevant stimuli in a practice run. This gave them an overall idea of the range of differences involved, helping them utilize the differential response scale presented during the actual experiments accordingly.

Data were analyzed using *Statistica 5.5* for Windows (StatSoft, Inc., 2000) and GANOVA-4, a custom statistical software incorporating the Linear Multinomial ANOVA model, part of Woodward et al. (1990).

The single non-perceptual experiment was designed and conducted using *Hypersignal's Block Diagram 4.2*, a digital signal processing development software by Hyperception, Inc. (2000).

6.2 Sound Wave Interference in Dichotic Experiments

6.2.1 Introduction

Reports of beating or beating-like sensations arising when two-component complex tones are presented dichotically (one in each ear) suggest that sound interference products may arise at a neural level. Such an interpretation questions the significance of a study addressing the energy content of amplitude fluctuations. Based on the theoretical arguments and pilot experiments discussed in Section 5.5 (regarding diplacusis, cross hearing, and directional cues) the present study claims that the perception of sound interference in terms of loudness changes, loudness fluctuations (beating), roughness, and/or combination tones cannot arise unless sound waves interact physically.

Three sets of experiments were conducted to challenge the claim of sound interference products arising due to neural wave interaction.

The first set of experiments (Section 6.2.2) examines one of the claim's untested predictions: beating Unisons due to diplacusis.

The second set of experiments (Section 6.2.3) examines whether changing the phase relationship between two identical sines presented dichotically will lead to observations compatible with the interference principle.

The third and final set of experiments (Section 6.2.4) examines whether dichotic beats, observed when two sines with slightly different frequencies are presented dichotically, are an indication of sound interference products arising at a neural level or a manifestation of a sound localization mechanism based on phase difference cues between the left and right ears (as suggested by the results of the preceding experiments). Section 6.2.5 summarizes the study results.

6.2.2 Experiment 1: Beating Unisons?

Hypothesis

Subjects demonstrating normal levels of diplacusis (frequency-pitch discrepancies between the two ears) will not report a beating/roughness sensation when presented binaurally with a sine stimulus, in spite of the fact that, at the neural level, there is a pitch difference between the left and right channels.

The above hypothesis statement is self-evident since, if that were not the case, everyone would hear beats when listening to Unisons. However, since the claimed possibility of interference at a neural level would require such beating, the hypothesis was tested experimentally.

Method

Ten subjects participated in this experiment. Prior to the experiment, a test was conducted to determine each subject's degree of diplacusis. Subjects were presented with the same sine signal through headphones, first in one ear and then the other (the order of presentation was randomized among subjects) and were asked to identify if there was any pitch difference between the two ears. The task was repeated

until the subjects felt confident of the pitch relationship between the two ears. They were then able to adjust the frequency of the tone fed to the ear reporting the higher pitch until they felt that the tones in both ears were the same in pitch. Frequency adjustments and tone synthesis were performed digitally using a computer program built for this purpose. Double-checking the results through amplifier and/or headphone channel reversal ensured that any reported pitch differences were not due to differences in the amplifier or headphone channels. Each subject repeated the task for four sine tones with frequencies 240Hz, 300Hz, 380Hz, and 480Hz, selected to span the range 200Hz-500Hz where dichotic beats have been most prominent in previous studies (i.e. Hornbostel, 1923, in von Békésy, 1960: 349; Plomp, 1966, 1976; Sethares, 1998; etc.).

In agreement with previous studies on sine-tone diplacusis patterns (see van den Brink, 1970), for the frequency values explored, all subjects reported a higher pitch for the right ear and a greater loudness for the left ear. This was also the case for the single left-handed subject, strengthening the argument that diplacusis has physiological rather than neurological roots (i.e. caused by minor differences between the two ears' basilar membrane elastic properties). This argument is supported further by the finding that the distribution of spontaneous otoacoustic emissions on the frequency axis correlates with the fine structure of the hearing threshold (i.e. Zwicker 1990; Zweig & Shera, 1995). Therefore, binaural diplacusis in normal ears seems to

not only depend on, but even originate from structural irregularities of the basilar membrane, supporting the *place theory*⁸⁵ of pitch.

Based on the frequency values obtained from the diplacusis tests, amplitude modulated sine signals were synthesized with the following characteristics: a) center frequencies equal to those used in the diplacusis exploration, b) modulation rates corresponding to the frequency difference necessary for equal pitches in both ears, and c) amplitude modulation (AM) depths 0%, 5%, 10%, and 25%. These signals served as comparison stimuli for the experiment. The justification for including the specified comparison stimuli was that, if interference products could arise at a neural level, subjects listening binaurally to sines through headphones should be observing beats at modulation rates matching their diplacusis levels and at a non-zero modulation depth.

For the actual experiment, the ten subjects listened through headphones to a total of sixty-eight pairs of stimuli (Table 6.1). They were asked to determine whether the tones in each pair were the same or different in terms of beating by rating each pair on a scale outlined by the labels: *Same beating* – *Different beating*. The response

⁸⁵ The *place theory* of pitch perception was developed by von Békésy (1963a&b) and suggests that, in the inner ear, different points on the basilar membrane undergo maximum displacement as a function of frequency. The neurons at the point of maximum displacement specify a particular frequency, based on their position on the basilar membrane. For a pure tone, the excitation pattern (i.e. the spatial distribution of mechanical or neural activity) along the basilar membrane is limited to one small, specified area (the critical band) with maximum excitation at a single position corresponding to the particular frequency. For a complex tone there is a pattern of excitations corresponding to the fundamental frequency and all the harmonics. Modifications to this theory are necessary in order to account for the phenomenon of the missing fundamental (or residue pitch) and for the pitch-shift effect (Section 5.4.2).

scale ranged from 0 (for same) to 100 (for different). Subjects were always comparing a monaural to a binaural stimulus to avoid combining any possible neural dichotic beats to any physically imposed beating. The only exemption was a single stimulus-pair per frequency group that compared the 0% AM-depth monaural stimulus to itself and was included as a control. The order of presentation (monaural-binaural / binaural-monaural) was randomized among subjects.

Table 6.1 displays the four groups of eight synthesized signals, one group for each of the four frequency values tested. For each group, all possible pairs between monaural and binaural stimuli were created along with a single control pair comparing the 0% AM-depth monaural stimulus to itself (a total of sixty-eight stimuli-pairs).

Table 6.1 List of Signals Used to Produce the Stimuli-Pairs for Experiment 1, Testing Possible Beating Due to Diplacusis There was a 250ms long pause inserted between the signals in each stimulus-pair. Modulation rates matched each subject's diplacusis levels.															
Signal group 1: 240Hz								Signal group 2: 300Hz							
<u>Monaural</u> %AM-depth				<u>Binaural</u> %AM-depth				<u>Monaural</u> %AM-depth				<u>Binaural</u> %AM-depth			
0	5	10	25	0	5	10	25	0	5	10	25	0	5	10	25
Signal group 3: 380Hz								Signal group 4: 480Hz							
<u>Monaural</u> %AM-depth				<u>Binaural</u> %AM-depth				<u>Monaural</u> %AM-depth				<u>Binaural</u> %AM-depth			
0	5	10	25	0	5	10	25	0	5	10	25	0	5	10	25

Results

Analysis of variance for one within-subjects random factor (frequency: 240Hz, 300Hz, 380Hz, 480Hz) was performed on the data revealing no frequency effect ($\alpha = 0.05$; $F(3,27) = 0.381$; $p = 0.767$). Data was therefore collapsed across the frequency factor and the mean responses of the subjects (17 means for the 17 stimuli pairs) are displayed in Table 6.2. This is an incomplete dissimilarity matrix. It results from all possible comparisons between monaural and binaural stimuli, and a single control comparison of the 0% AM-depth monaural stimulus to itself.

Table 6.2
Mean Responses (10 Subjects) for the 17 Stimuli-Pair Categories in Experiment 1,
Collapsed across the Frequency Factor

Subjects rated the second stimulus in each pair relative to the first in terms of the degree of beating, on a scale outlined by the labels: *Same beating - Different beating*. The response scale ranged from 0 (for same) to 100 (for different). Each cell represents the degree of similarity/dissimilarity between a monaural and a binaural stimulus in terms of their degree of beating. The first value in the split cell represents the observed similarity/dissimilarity of the 0% AM-depth monaural stimulus to itself.

<u>Binaural</u> <u>AM depth</u>	0%		5%	10%	25%
<u>Monaural</u> <u>AM depth</u>					
0%	3.425	4.325	24.75	66.25	95.7
5%	26.225		5.525	30.975	81.35
10%	66.875		32.8	5.45	38.125
25%	95.3		78.65	37.725	5.025

Analysis of variance (Scheffé test; $\alpha = 0.05$) was performed on the means to examine:

- whether there is a significant difference in the degree of beating for monaural and binaural stimuli of the same modulation depth (means on the diagonal of Table 6.2) and
- what modulation depth in the monaural stimuli matches in beating the binaural stimulus with 0% modulation depth (means on the first column of Table 6.2).

The results are displayed in Figures 6.1 & 6.2.

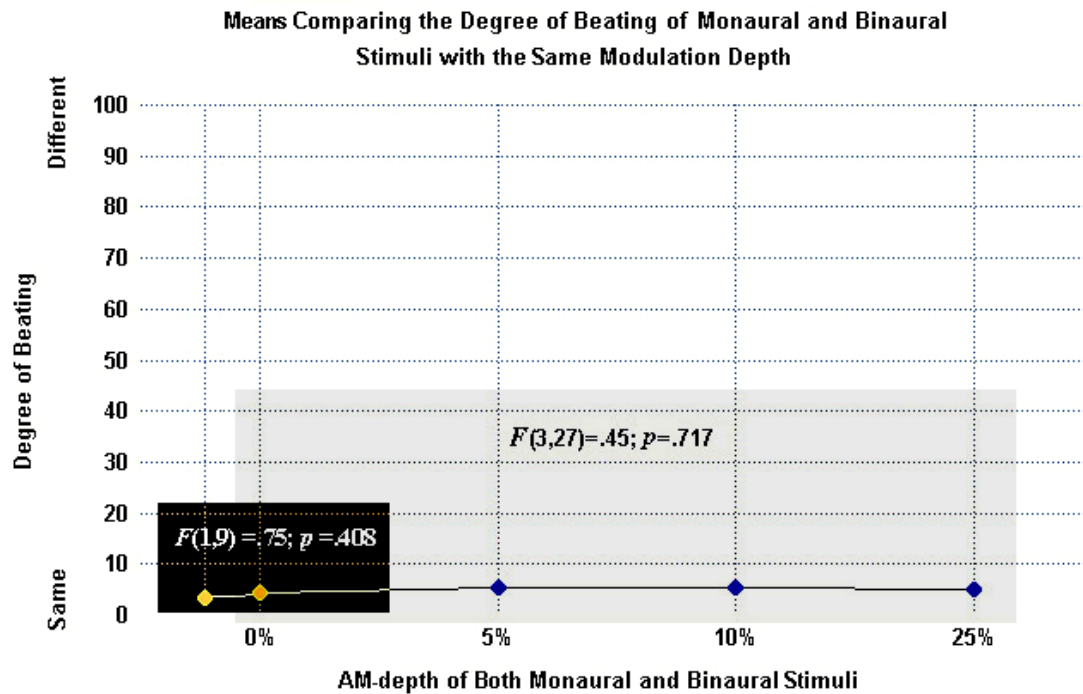


Figure 6.1: Mean responses (10 subjects) comparing the degree of beating of monaural and binaural stimuli with the same AM-depth (four means in the grayed area), and the degree of beating of the monaural stimulus with AM-depth = 0% to itself (first mean in the blackened area); Experiment 1.

Figure 6.1 (grayed area) indicates that subjects reported the same degree of beating for monaural and binaural stimuli with the same AM-depth ($F(3,27) = 0.45$; $p = 0.717$). In other words, the sameness / difference in the degree of beating for monaural and binaural stimuli with the same modulation depth does not change, regardless of modulation depth. More importantly, as indicated by the two means in the blackened area of the graph, the degree of beating for AM-depth = 0% is the same regardless of whether a sine is presented monaurally or binaurally (there was no significant difference between the means comparing a) the beating of the monaural and binaural stimuli with AM-depth = 0% (mean = 4.325) and b) the beating of the monaural stimulus with AM-depth = 0% to itself (mean = 3.425)) ($F(1,9) = 0.75$ $p = 0.408$).

Figure 6.2 (next page) demonstrates that subjects reported significant differences between the degree of beating of the binaural stimulus with AM-depth = 0% and the degree of beating of the monaural stimuli with AM-depths >0% ($F(3,27) = 1481$; $p < 0.001$).

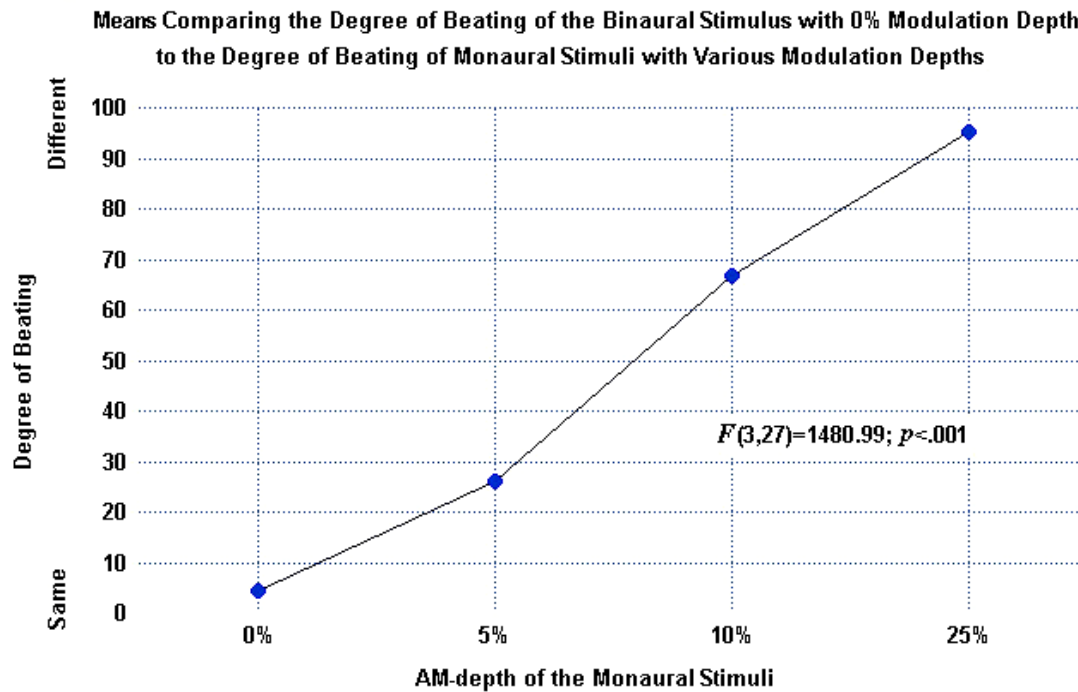


Figure 6.2: Mean responses (10 subjects) comparing the degree of beating of the binaural stimulus with AM-depth = 0% to the degree of beating of monaural stimuli with various AM-depths. Experiment 1.

Figures 6.1 & 6.2 demonstrate that subjects report the same degree of beating for amplitude modulated sines with the same modulation depth, regardless of whether the stimuli are presented monaurally or binaurally. In other words, although for binaural sine stimuli there is a pitch difference between the two ears that would introduce additional beating if interference products could arise at a neural level, no such additional beating is observed. Therefore, the results of Experiment 1 do not support the claim of sound interference products arising due to neural wave interaction.

6.2.3 Experiments 2a, b, & c: Dichotic Interference / Cancellation?

Hypothesis

Changing the phase relationship between two otherwise identical sines presented in stereo (one sine in each channel) through headphones, will lead to observations incompatible with the interference principle:

- a) No loudness change will be reported.*
- b) For frequencies below 500Hz, phase changes between the two ears will influence the perceived direction of the two-component stimuli, with 180° out of phase stimuli having a wider stereo image than in phase stimuli.*
- c) For frequencies above 1500Hz, changing the phase relationship between the two ears will result in no perceptual change.*

The above hypothesis is based on the discussion in Section 5.5 and investigates the phase-based sound localization mechanism hypothesized in the related literature, linking phase changes between ears to directional rather than loudness changes.

Method

Three experiments were conducted with a total of thirty subjects. For each experiment, ten subjects listened through headphones to the stimuli-pairs described in the following Sections. Each subject was asked to rate the second stimulus in each pair relative to the first stimulus in terms of:

Experiment 2a: Loudness. Subjects rated the second stimulus relative to the first on a scale outlined by the labels: *Softer - Louder*.

Experiment 2b: Direction. Subjects rated the second stimulus relative to the first on a scale outlined by the labels: *More to the left - More to the right*.

Experiment 2c: Stereo image width. Subjects rated the second stimulus relative to the first on a scale outlined by the labels: *Narrower - Wider*.

For all experiments, subjects entered their responses by dragging a scroll-bar on a computer screen along an appropriately marked scale (the scroll-bar starting position was fixed at the center).

6.2.3.1 Experiment 2a: Effect of Dichotic Phase Differences on Loudness

Subjects listened to two groups of stereo stimuli-pairs. The frequency of the stimuli was 300Hz in the first group and 1700Hz in the second group. These frequency values were selected to represent frequency ranges below 500Hz and above 1500Hz respectively, ranges where the relationship between phase changes in the sound arriving at the two ears and sound localization cues is unambiguous. The stimuli within each group were made out of two sines (one in the left and one in the right channel) with the same frequency, same amplitude, and various phase relationships between channels.

The stimuli lists, shown in Table 6.3 (next page), also included two stimuli per group with the channels in-phase, the same frequency as the rest of the stimuli in the group, and intensity reduced relative to the other stimuli by 2.5dB and 5dB respectively. These stimuli represented actual loudness changes, which would result from physical interference between two identical sines with 90^0 and 120^0 phase difference, and were included as controls. All possible stimuli-pairs within each group were created, resulting in a total of seventy-two ($2 \times (6 \times 6)$) stimuli-pairs.

Table 6.3 List of Stimuli Used to Produce the Stimuli-Pairs for Experiment 2a, Testing the Effect of Dichotic Phase Differences on Loudness A 250ms long pause was inserted between the signals in each stimulus-pair.	
Signal group 1: 300Hz	Signal group 2: 1700Hz
Stereo - In phase: Left 0° Right 0°	Stereo - In phase: Left 0° Right 0°
Stereo - Out of phase: Left 90° Right 0°	Stereo - Out of phase: Left 90° Right 0°
Stereo - Out of phase: Left 0° Right 90°	Stereo - Out of phase: Left 0° Right 90°
Stereo - Out of phase: Left 180° Right 0°	Stereo - Out of phase: Left 180° Right 0°
Stereo - In Phase (-2.5dB)	Stereo - In Phase (-2.5dB)
Stereo - In Phase (-5dB)	Stereo - In Phase (-5dB)

The order of presentation (group 1-group 2 / group 2-group 1) was randomized among subjects. Ten subjects listened through headphones to the seventy-two stimuli-pairs. Each subject was asked to rate the second stimulus in each pair relative to the first stimulus in terms of loudness, on a scale outlined by the labels: *Softer - Louder* and ranging from 0 (softer) to 100 (louder) (50: same).

Results

Analysis of variance for one within-subjects random factor (frequency: 300Hz, 1700Hz) was performed on the data revealing no frequency effect ($\alpha = 0.05$; $F(1,9) = 0.75$; $p = 0.79$). Data was collapsed across the frequency factor and the mean responses of the subjects (36 means for the 36 stimuli-pair categories) are displayed in

Table 6.4. This is a dissimilarity matrix resulting from all possible comparisons among the six stimuli described in either column of Table 6.3, above.

Table 6.4 Mean Responses (10 Subjects) for 36 Stimuli-Pair Categories in Experiment 2a, Collapsed Across the Frequency Factor Subjects rated the second stimulus in each pair relative to the first in terms of loudness, on a scale outlined by the labels: <i>Softer - Louder</i> and ranging from 0 (softer) to 100 (louder) (50: same). Each cell represents the degree of similarity/dissimilarity between the stimuli in each pair in terms of loudness.						
<u>Stimuli Pairs</u>	In Phase L: 0 ⁰ R: 0 ⁰ +0dB	Out of Phase L: 90 ⁰ R: 0 ⁰ +0dB	Out of Phase L: 0 ⁰ R: 90 ⁰ +0dB	Out of Phase L: 180 ⁰ R: 0 ⁰ +0dB	In Phase L: 0 ⁰ R: 0 ⁰ -2.5dB	In Phase L: 0 ⁰ R: 0 ⁰ -5dB
In Phase L: 0 ⁰ R: 0 ⁰ +0dB	48.450	49.000	49.100	47.850	28.400	3.550
Out of Phase L: 90 ⁰ R: 0 ⁰ +0dB	48.400	48.950	47.950	49.600	28.250	2.900
Out of Phase L: 0 ⁰ R: 90 ⁰ +0dB	49.850	49.050	48.500	49.000	27.550	3.200
Out of Phase L: 180 ⁰ R: 0 ⁰ +0dB	49.300	48.650	48.900	49.100	28.100	3.750
In Phase L: 0 ⁰ R: 0 ⁰ -2.5dB	73.000	73.650	73.100	74.250	49.100	25.650
In Phase L: 0 ⁰ R: 0 ⁰ -5dB	94.300	95.400	95.700	96.600	73.850	49.550

The means in the dissimilarity matrix (Table 6.4) were subjected to cluster analysis⁸⁶ and the results are displayed in Figure 6.3.

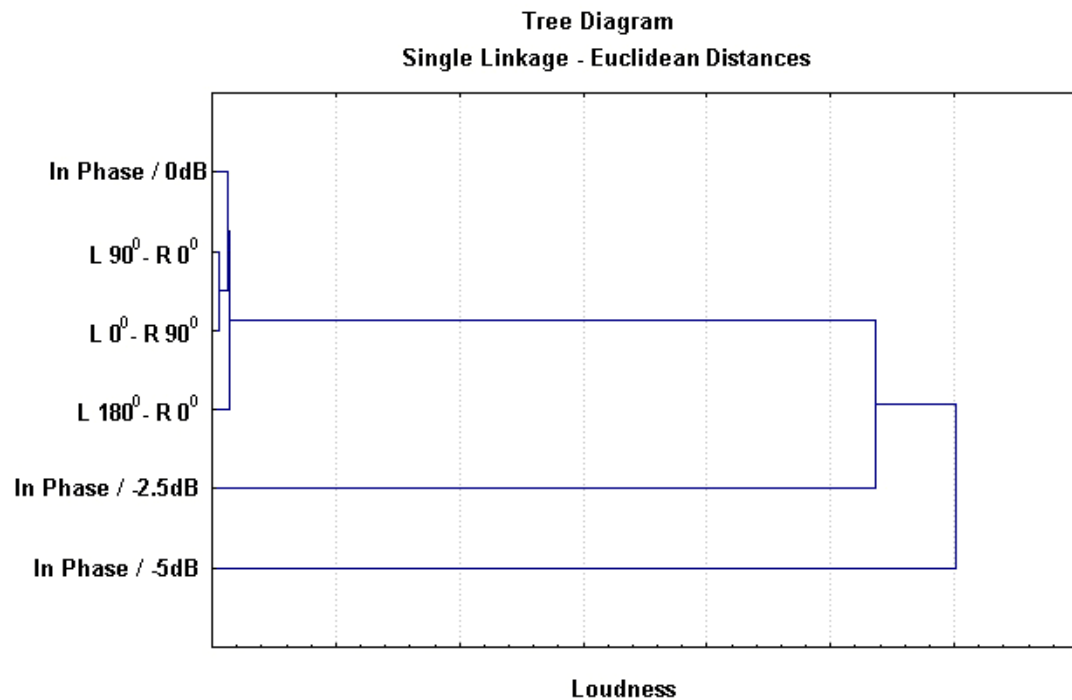


Figure 6.3: Tree diagram resulting from cluster analysis of the means (10 subjects; Table 6.4) in Experiment 2a. The first four stimuli categories have been grouped in a single loudness cluster, indicating that changing the phase relationship between the left and right channels of a sinusoidal stimulus presented in stereo does not result in a loudness change.

In Figure 6.3, the first four stimuli categories have been grouped in a single loudness cluster, separated from the stimuli categories that included an intensity reduction. Analysis of variance (Scheffé test; $\alpha = 0.05$) confirmed that there was a

⁸⁶ The data means represent distances and were used raw for the analysis (no distance matrix had to be computed). The *Single Linkage* method (i.e. using the 'nearest neighbors' across clusters) was

significant difference in the mean responses for the in-phase stimuli with different intensities ($F(2,18) = 542.44$; $p < 0.001$), while there was no significant difference in the mean responses for the stimuli with equal intensities and variable phase between the left and right ears ($F(3,27) = 1.126$, $p = 0.356$).

The results demonstrate that, changing the phase relationship between the left and right channels of a sinusoidal stimulus presented in stereo through headphones does not influence its loudness. If interference products could arise at a neural level, phase changes in the above experiment would have resulted in loudness changes. The absence of such changes supports the claim against sound interference products arising due to neural wave interaction.

used to determine the distances among clusters. This method was chosen because, prior to the analysis, the assumption that the data means would form distinct 'clumps' was not justified.

6.2.3.2 Experiment 2b: Effect of Dichotic Phase Differences on Direction

The stimuli used to create the stimuli-pairs for this experiment were the same as those in Experiment 2a with the following exemption: The four control stimuli were replaced by four in-phase stimuli that reflected changes in direction rather than loudness. Each control stimulus had the same frequency as the rest of the stimuli in the group and was panned to the left or to the right as indicated in Table 6.5 (next page). Panning was simulated through an intensity difference between the left and right channels (8dB), chosen to match the apparent panning of the out of phase stimuli. Again all possible stimuli pairs within each group were created, resulting in a total of seventy-two (2x(6x6)) stimuli-pairs. The order of presentation (group 1-group 2 / group 2-group 1) was randomized among subjects.

<p>Table 6.5 List of Stimuli Used to Produce the Stimuli-Pairs for Experiment 2b, Testing the Effect of Dichotic Phase Differences on Direction</p> <p>All possible pairs within each group were created, resulting in a total of $2 \times (6 \times 6) = 72$ stimuli-pairs. A 250ms long pause was inserted between the signals in each stimulus-pair. Stimuli-pairs were presented in stereo.</p>	
Signal group 1: 300Hz	Signal group 2: 1700Hz
Stereo - In phase: Left 0^0 Right 0^0	Stereo - In phase: Left 0^0 Right 0^0
Stereo – Out of phase: Left 90^0 Right 0^0	Stereo - Out of phase: Left 90^0 Right 0^0
Stereo – Out of phase: Left 0^0 Right 90^0	Stereo - Out of phase: Left 0^0 Right 90^0
Stereo - Out of phase: Left 180^0 Right 0^0	Stereo - Out of phase: Left 180^0 Right 0^0
Stereo - In phase Panned to the Left (L-R = 8dB)	Stereo - In phase Panned to the Left (L-R = 8dB)
Stereo - In phase Panned to the Right (R-L = 8dB)	Stereo - In phase Panned to the Right (R-L = 8dB)

Ten subjects listened through headphones to the seventy-two stimuli-pairs. Each subject was asked to rate the second stimulus in each pair relative to the first stimulus in terms of direction, on a scale outlined by the labels: *More to the left* - *More to the right* and ranging from 0 (more to the left) to 100 (more to the right) (50: same).

Results

Analysis of variance for one within-subjects random factor (frequency: 300Hz, 1700Hz) was performed on the data. There was a clear frequency effect ($\alpha = 0.05$; $F(1,9) = 186.16$; $p < 0.001$), indicating that phase changes influenced the perceived

direction of the stimuli differently for the 300Hz and the 1700Hz stimuli groups. Data means are displayed in Tables 6.6 (300Hz) and 6.7 (1700Hz). These are dissimilarity matrixes, resulting from all possible comparisons among the six stimuli described in each one of the columns of Table 6.5, above.

Table 6.6 Mean Responses (10 Subjects) for 36 Stimuli-Pairs in Experiment 2b (300Hz) Subjects rated the second stimulus in each pair relative to the first in terms of direction, on a scale outlined by the labels: <i>More to the Left - More to the Right</i> and ranging from 0 (more to the left) to 100 (more to the right) (50: same). Each cell represents the degree of similarity/dissimilarity between the stimuli in each pair in terms of direction.						
<u>Stimuli Pairs</u>	In Phase L: 0 ⁰ R: 0 ⁰ L-R=0dB	Out of Phase L: 90 ⁰ R: 0 ⁰ L-R=0dB	Out of Phase L: 0 ⁰ R: 90 ⁰ L-R=0dB	Out of Phase L: 180 ⁰ R: 0 ⁰ L-R=0dB	In Phase L: 0 ⁰ R: 0 ⁰ L-R=8dB	In Phase L: 0 ⁰ R: 0 ⁰ R-L=8dB
In Phase L: 0 ⁰ R: 0 ⁰ L-R=0dB	47.7	31.8	68.1	48.6	23.2	75.6
Out of Phase L: 90 ⁰ R: 0 ⁰ L-R=0dB	62.7	48.5	79.2	63.6	44.4	86.1
Out of Phase L: 0 ⁰ R: 90 ⁰ L-R=0dB	37.1	18.2	47.5	37	16.4	52.2
Out of Phase L: 180 ⁰ R: 0 ⁰ L-R=0dB	48.3	34.6	63.4	49.5	29.1	78.5
In Phase L: 0 ⁰ R: 0 ⁰ L-R=8dB	65.6	52.3	90.2	66.3	49.5	93.6
In Phase L: 0 ⁰ R: 0 ⁰ R-L=8dB	32.4	8.8	46.5	33.9	7.3	48.5

The means in the dissimilarity matrix (Table 6.6) were subjected to cluster analysis and the results are displayed in Figure 6.4.

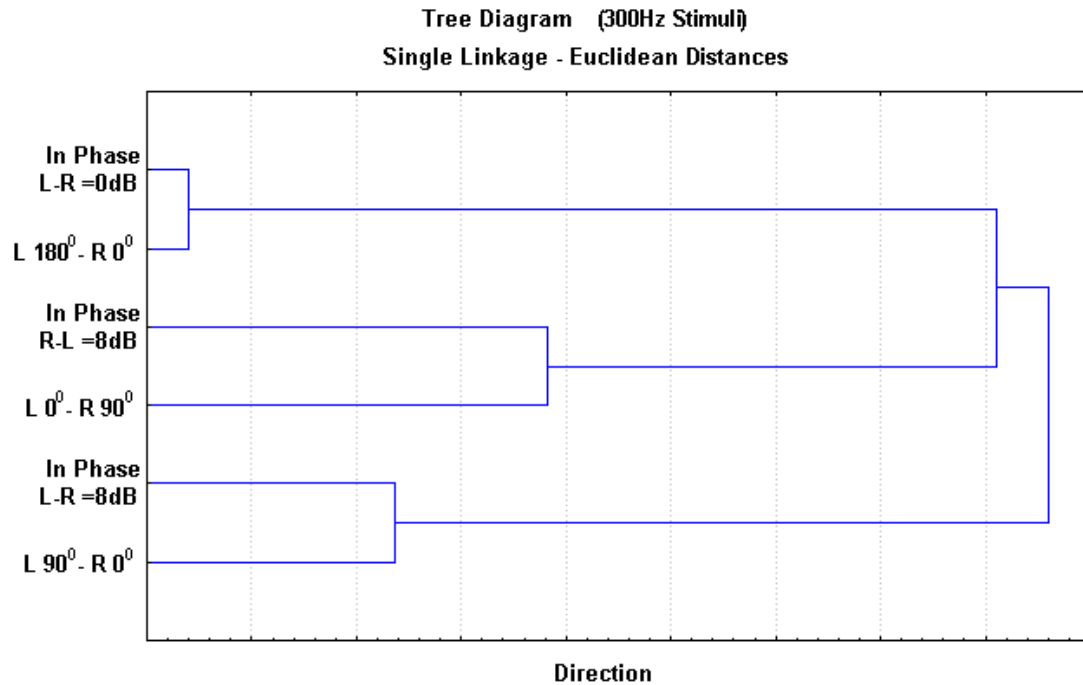


Figure 6.4: Tree diagram resulting from cluster analysis of the means (10 subjects; Table 6.6, 300Hz) in Experiment 2b. Three direction clusters can be identified with two stimuli each.

Three separate clusters can be identified in Figure 6.4, with two stimuli each. This grouping was confirmed through analysis of variance (Scheffé test; $\alpha = 0.05$): One out of phase stimulus (L: 90° R: 0°) was grouped with the in phase stimulus panned to the left (In Phase, L-R = 8dB) ($F(1,9) = 2.14$; $p = 0.177$). The second out of phase stimulus (L: 0° R: 90°) was grouped with the in phase stimulus panned to the

right (In Phase, R-L = 8dB) ($F(1,9) = 2.307; p = 0.163$). The last out of phase stimulus (L: 180^0 R: 0^0) was grouped with the in phase stimulus panned in the middle (In Phase, L-R = 0dB) ($F(1,9) = 0.669; p = 0.434$).

It is evident that the six 300Hz stimuli have been grouped according to panning/direction into three different *direction* groups ($F(2,18) = 211.5; p < 0.001$). This demonstrates that changing the phase relationship between the left and right channels of a 300Hz sinusoidal stimulus presented in stereo through headphones affects its perceived direction (180^0 out of phase stimuli are grouped in the same direction cluster as the in phase stimuli).

This is not the case for the 1700Hz stimuli. Table 6.7 (next page) displays the data means for the 1700Hz stimulus group in a dissimilarity matrix.

<p>Table 6.7 Mean Responses (10 Subjects) for 36 Stimuli Pairs in Experiment 2b (1700Hz) Subjects rated the second stimulus in each pair relative to the first in terms of direction, on a scale outlined by the labels: <i>More to the Left - More to the Right</i> and ranging from 0 (more to the left) to 100 (more to the right) (50: same). Each cell represents the degree of similarity/dissimilarity between the stimuli in each pair in terms of direction.</p>						
<u>Stimuli Pairs</u>	In Phase L: 0 ⁰ R: 0 ⁰ L-R=0dB	Out of Phase L: 90 ⁰ R: 0 ⁰ L-R=0dB	Out of Phase L: 0 ⁰ R: 90 ⁰ L-R=0dB	Out of Phase L: 180 ⁰ R: 0 ⁰ L-R=0dB	In Phase L: 0 ⁰ R: 0 ⁰ L-R=8dB	In Phase L: 0 ⁰ R: 0 ⁰ R-L=8dB
In Phase L: 0 ⁰ R: 0 ⁰ L-R=0dB	47.4	49.7	48.5	48.4	26	75.7
Out of Phase L: 90 ⁰ R: 0 ⁰ L-R=0dB	49.7	48.5	48.3	50.3	26.6	76.6
Out of Phase L: 0 ⁰ R: 90 ⁰ L-R=0dB	49.7	48.6	48.7	48.1	25.4	75.5
Out of Phase L: 180 ⁰ R: 0 ⁰ L-R=0dB	48.4	48.5	50	48.1	24.4	74.4
In Phase L: 0 ⁰ R: 0 ⁰ L-R=8dB	73.3	72.3	71.8	71.7	49.3	94.3
In Phase L: 0 ⁰ R: 0 ⁰ R-L=8dB	25.9	26.3	27.2	25.7	2.5	49.7

Data means were subjected to cluster analysis and the results are displayed in Figure 6.5.

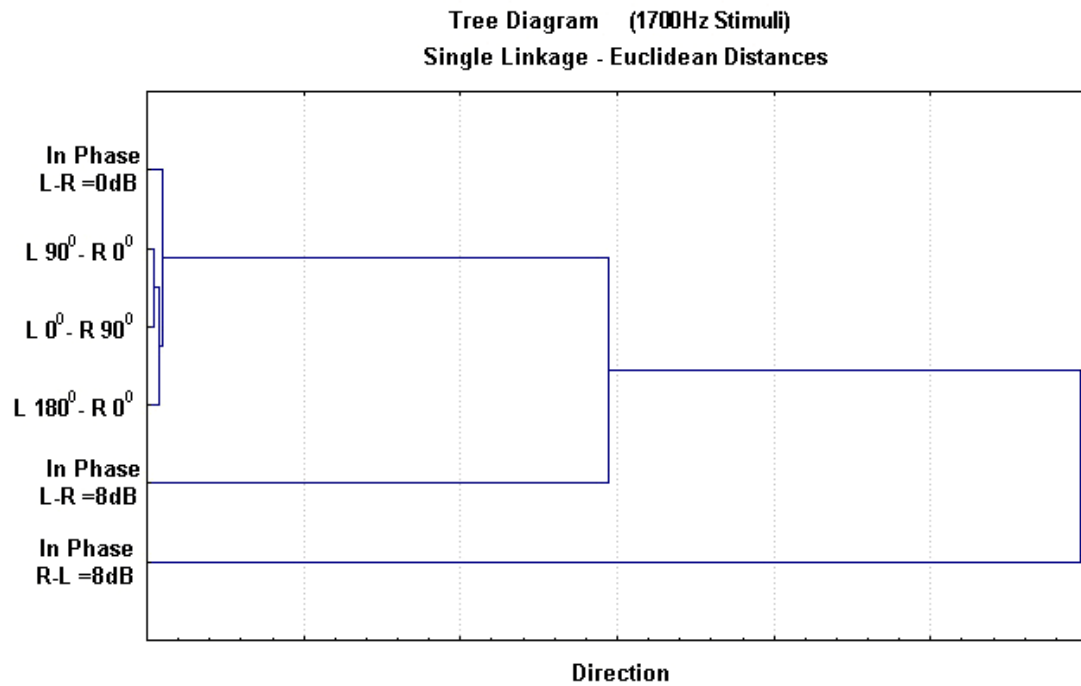


Figure 6.5: Tree diagram resulting from cluster analysis of the means (10 subjects; Table 6.7, 1700Hz) in Experiment 2b. The first four stimuli have been grouped in a single *direction* cluster.

In Figure 6.5, the first four stimuli have been grouped in a single *direction* cluster, separated from the stimuli that included a direction change. Analysis of variance (Scheffé test; $\alpha = 0.05$) confirmed that there was a significant difference in the mean responses for the in-phase stimuli with different panning ($F(2,18) = 200.41$; $p < 0.001$), while there was no significant difference in the mean responses for the

stimuli with the same panning and variable phase between the left and right ears ($F(3,27) = 1.613, p = 0.209$). The results demonstrate that, changing the phase relationship between the left and right channels of a 1700Hz sinusoidal stimulus presented in stereo through headphones does not change the perceived direction of the stimulus.

The results of Experiment 2b are inconsistent with sound interference expectations and in agreement with studies examining sound localization in the free field. These studies indicate that, for frequencies $< 500\text{Hz}$ (i.e. 300Hz), the most important directional cues on the horizontal plane come from signal phase-differences between the two ears (Hartmann, 1999: 24). For frequencies $> 1500\text{Hz}$ (i.e. 1700Hz) this mechanism performs poorly (Hartmann, 1999: 25). Similarly, Experiment 2b has demonstrated that changing the phase relationship between the left and right channels of a sinusoidal stimulus presented in stereo through headphones affects its perceived direction at 300Hz but not at 1700Hz.

6.2.3.3 Experiment 2c: Effect of Dichotic Phase Differences on Stereo Image Width

For this experiment, eight stereo stimuli were synthesized (Table 6.8). Each stimulus was made out of two sines (one per channel) with the same frequency and amplitude values, and 0^0 or 180^0 phase difference between the left and right channels (i.e. *in phase* and *180⁰ out of phase*). Stimuli were synthesized at 200Hz and 300Hz (to represent frequencies <500Hz) as well as 1700Hz and 2500Hz (to represent frequencies >1500Hz). All possible stimuli pairs were created resulting in a total of sixty-four stimuli-pairs.

Table 6.8 List of Stimuli Used to Produce the Stimuli-Pairs for Experiment 2c, Testing the Effect of Dichotic Phase Differences on Stereo Image Width All possible pairs were created, resulting in a total of 8x8=64 stimuli-pairs. A 250ms long pause was inserted between the signals in each stimulus-pair.							
< 500Hz				> 1500Hz			
200Hz		300Hz		1700Hz		2500Hz	
<u>In Phase</u> L: 0^0 R: 0^0	<u>Out of Phase</u> L: 180^0 R: 0^0	<u>In Phase</u> L: 0^0 R: 0^0	<u>Out of Phase</u> L: 180^0 R: 0^0	<u>In Phase</u> L: 0^0 R: 0^0	<u>Out of Phase</u> L: 180^0 R: 0^0	<u>In Phase</u> L: 0^0 R: 0^0	<u>Out of Phase</u> L: 180^0 R: 0^0

Ten subjects listened through headphones to the sixty-four stimuli-pairs. Each subject was asked to rate the second stimulus in each pair relative to the first stimulus in terms of stereo image width, on a scale outlined by the labels: *Narrower - Wider*, and ranging from 0 (narrower) to 100 (wider) (50: same).

Results

Table 6.9 is a dissimilarity matrix displaying the data means from all possible comparisons among the eight stimuli described in Table 6.8.

Table 6.9 Mean Responses (10 Subjects) for 64 Stimuli-Pairs in Experiment 2c Subjects rated the second stimulus in each pair relative to the first in terms of stereo image width. Responses were entered using a scale outlined by the labels: <i>Narrower - Wider</i> and ranging from 0 (narrower) to 100 (wider) (50: same). Each cell represents the degree of similarity/dissimilarity between the stimuli in each pair in terms of stereo image width.								
<u>Stimuli Pairs</u>	200Hz L: 0 ⁰ R: 0 ⁰	200Hz L: 180 ⁰ R: 0 ⁰	300Hz L: 0 ⁰ R: 0 ⁰	300Hz L: 180 ⁰ R: 0 ⁰	1700Hz L: 0 ⁰ R: 0 ⁰	1700Hz L: 180 ⁰ R: 0 ⁰	2500Hz L: 0 ⁰ R: 0 ⁰	2500Hz L: 180 ⁰ R: 0 ⁰
200Hz L: 0 ⁰ R: 0 ⁰	47.7	99.2	48.6	98	48.5	47.2	48.3	49.5
200Hz L: 180 ⁰ R: 0 ⁰	2.7	48.5	1.9	49.5	1.8	1.4	2	2.7
300Hz L: 0 ⁰ R: 0 ⁰	50	97.5	49.7	98.4	48.6	49.5	49.7	48.5
300Hz L: 180 ⁰ R: 0 ⁰	2.4	48.6	1.1	50	2.6	1.5	1.7	2.2
1700Hz L: 0 ⁰ R: 0 ⁰	48.5	98.3	49.8	98	49.3	48.6	48.7	48.1
1700Hz L: 180 ⁰ R: 0 ⁰	49.7	97.7	49.7	96.9	50	48.1	48.4	48.5
2500Hz L: 0 ⁰ R: 0 ⁰	49.7	98.1	47.4	98	48.5	48.4	50.3	49.7
2500Hz L: 180 ⁰ R: 0 ⁰	48.5	97.9	48.3	98.5	47.7	48.5	48.6	47.2

Analysis of variance was performed on the data ($\alpha = 0.05$) for 3 within-subjects factors:

- a) *Frequency Range* (fixed factor with 2 levels: Under 500Hz / Over 1500Hz)
- b) *Frequency* (random factor with two levels nested under *Frequency Range*: 200Hz - 300Hz / 1700Hz - 2500Hz) and
- c) *Phase* (random factor with two levels nested under *Frequency*: In Phase / 180° Out of Phase.)

The analysis indicated no effect of frequency within each frequency range level (<500Hz: $F(1,9) = 0.192$; $p = 0.671$ - >1500Hz: $F(1,9) = 2.567$; $p = 0.144$). Data was therefore collapsed accordingly and the 16 data means for 10 subjects are displayed in Table 6.10 (next page). This is a dissimilarity matrix, comparing all possible combinations of the frequency range and phase factors.

Table 6.10 Mean Responses (10 Subjects) Comparing All Combinations between the Frequency Range (< 500Hz / > 1700Hz) and Phase (In Phase / 180° Out of Phase) Factors in Experiment 2c Responses were entered on a scale outlined by the labels: <i>Narrower - Wider</i> and ranging from 0 (narrower) to 100 (wider) (50: same). Each cell represents the degree of similarity/dissimilarity between the stimuli categories in each stimulus pair in terms of stereo image width.				
<u>Stimuli Pairs</u>	< 500Hz In Phase L: 0° R: 0°	< 500Hz Out of Phase L: 180° R: 0°	> 1500Hz In Phase L: 0° R: 0°	> 1500Hz Out of Phase L: 180° R: 0°
< 500Hz In Phase L: 0° R: 0°	49	98.275	48.775	48.675
< 500Hz Out of Phase L: 180° R: 0°	2.025	49.15	2.025	1.95
> 1500Hz In Phase L: 0° R: 0°	48.85	98.1	49.2	48.7
> 1500Hz Out of Phase L: 180° R: 0°	49.05	97.75	48.675	48.075

Data means were subjected to cluster analysis⁸⁷ and the results are displayed in Figure 6.6.

⁸⁷ The *Complete Linkage* method (using the 'furthest neighbors' across clusters) was used to determine the distances between clusters. This method performs well in cases where the means are expected to form naturally distinct 'clumps'. Such an expectation was justified in Experiment 2c since only two extreme phase values were used.

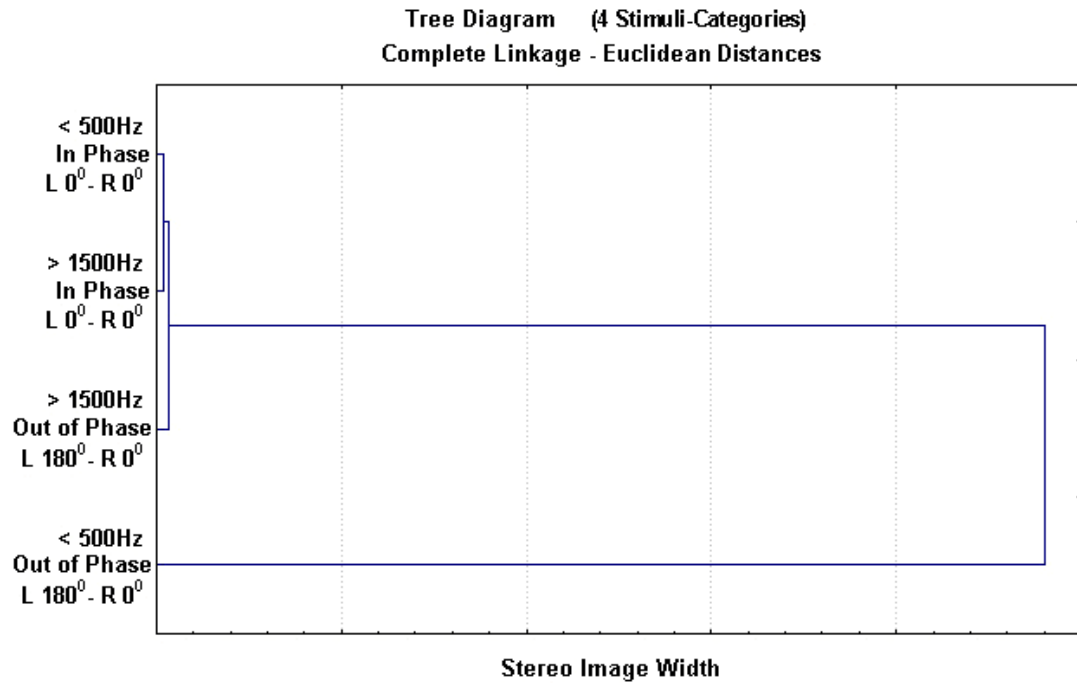


Figure 6.6: Tree diagram resulting from cluster analysis of the means (10 subjects; Table 6.10) in Experiment 2c. Two clusters can be identified. A 'narrow stereo image' cluster containing three categories of stimuli and a 'wide stereo image' cluster containing one stimulus category.

In Figure 6.6, stimuli categories have been grouped under two distinct clusters. The first includes the *in phase* stimuli with frequencies 200Hz and 300Hz ($<500\text{Hz}$), along with the *in phase* and the 180° *out of phase* stimuli with frequencies 1700Hz and 2500Hz ($>1500\text{Hz}$). Based on the experimental question, this cluster can be interpreted as a 'narrow stereo image' cluster. The second cluster includes only the 180° *out of phase* stimuli with frequencies 200Hz and 300Hz ($<500\text{Hz}$) and can be interpreted as a 'wide stereo image cluster'. Analysis of variance (Sheffé test, $\alpha = 0.05$) confirmed the cluster analysis results. There was no significant difference in the mean stereo image width of the stimuli categories in the first cluster ($F(2,18) = 0.254$;

$p = 0.779$), while there was a significant difference between the mean stereo image widths of the stimuli in the first and second clusters ($F(1,9) = 9734.86$; $p < 0.001$).

The analysis demonstrates that, for frequencies $< 500\text{Hz}$ (200Hz & 300Hz), 180° phase difference between the left and right channels of dichotically presented sine stimuli (L: 180° R: 0°) results in a wider stereo image than 0° phase difference (L: 0° R: 0°). For frequencies $> 1500\text{Hz}$ (1700Hz & 2500Hz), 180° phase difference between the left and right channels has no effect on the stereo image width of the sound.

As was the case with Experiment 2b, the results of Experiment 2c are inconsistent with sound interference expectations and in agreement with studies on sound localization. Combined with the results from Experiments 2a & 2b, they demonstrate that phase changes between the left and right channels of dichotic sine stimuli have directional, rather than loudness implications for frequencies $< 500\text{Hz}$ (200Hz & 300Hz) and no implications for frequencies $> 1500\text{Hz}$ (1700Hz & 2500Hz).

6.2.3.4 Discussion and Implications of Experiments 2a, b, & c

Experiments 2a, 2b, & 2c were designed to test whether or not changes in the phase relationship between otherwise identical sine signals presented dichotically would give rise to perceptions compatible with the interference principle. If interference products could arise from neural wave interaction, phase changes between the two channels of dichotically presented stereo sine stimuli should have given rise to loudness changes. However, no loudness changes were reported. Instead, phase changes had directional rather than loudness implications:

- a) For frequencies <500Hz (200Hz & 300Hz), phase changes between the two ears influenced the perceived direction of the stimuli, with 180° out of phase stimuli having a wider stereo image than in phase stimuli.
- b) For frequencies >1500Hz (1700Hz & 2500Hz), phase changes between the two ears resulted in no perceptual change.

The results of Experiments 2a, 2b, & 2c are inconsistent with interference expectations and consistent with results from sound localization studies (Hartmann, 1999). They can help to better understand why subjects listening dichotically to two sines with slightly different frequencies (one sine per ear) have reported beating-like sensations. The experimental results support the hypothesis that, in dichotic explorations of beats the constantly shifting phase between the sines presented in the

two ears results in a constantly shifting localization of the combined sound. For low frequency differences, this shift is perceived as a rotation of the sound inside the subject's head and, for higher frequency differences, it results in what has been termed loudness or timbral fluctuation.⁸⁸ At frequency differences $< 3\text{Hz}$, the sensation of sound rotation is clearly identifiable, while for frequency differences $> 3\text{Hz}$, the sensation becomes increasingly confused with beating (von Békésy, 1960).

The discussion indicates that the reported beating-like sensation may have been the manifestation of a sound localization mechanism based on phase difference cues rather than an indication of neural wave interaction giving rise to sound interference products. Such an explanation is further confirmed by the results of pilot experiments where a sine tone was presented to subjects through headphones and made to move/rotate between the left and right ears (pan-pot effect). For $\approx \leq 3$ rotations per second, subjects perceive the sound moving from one ear to the other. For $\approx \geq 3$ rotations per second, however, subjects were not aware of the rotating motion but reported a beating-like sensation at a rate corresponding to the rotation rate.⁸⁹

⁸⁸ This observation is consistent with reports from numerous studies (i.e. Hornbostel, 1923, in von Békésy, 1960: 349; Lane, 1925; von Békésy, 1960; Plomp, 1966; Hartmann, 1999; etc.).

⁸⁹ Lane (1925) conducted a similar experiment using a rotating sound source in an anechoic chamber.

As confirmed through Experiments 2b & 2c, the phase-based sound localization mechanism works poorly for frequencies beyond 1500Hz (1700Hz & 2500Hz). It should therefore be possible to test the above explanation using stimuli at low levels and high frequencies. To date, very few studies examining dichotic beats have used stimuli with frequencies over 1500Hz (i.e. Lane, 1925). In a preliminary experiment, a 2000Hz sine tone was presented to the left ear and sine tones between 2001Hz and 2010Hz were presented to the right (at 1Hz intervals). No beating or beating-like sensation was reported, providing further evidence against the hypothesis of sound interference products arising at a neural level. The following experiment (Experiment 3) is based on the results of Experiments 2a, 2b, & 2c and reexamines the possibility of dichotic beats, taking into account issues of sound localization cues.

6.2.4 Experiment 3: Dichotic Beats and Sound Localization Cues

Hypothesis

Whenever dichotic beats are observed, they are a manifestation of a sound localization mechanism based on phase difference cues between the left and right ears rather than an indication of sound interference products arising at a neural level.

Method

Ten subjects listened to two groups of stereo stimuli, one centered at 300Hz and one centered at 1700Hz. These frequency values were selected to represent frequency ranges below 500Hz and above 1500Hz respectively, ranges where the relationship between phase changes in the sound arriving at the two ears and sound localization cues is unambiguous. All stimuli had two components (one in the left and one in the right channel) with the same amplitudes and phases and the frequency relationships shown in the stimuli list (Table 6.11). The stimuli list included the following comparison stimuli: a) amplitude-modulated (AM) sines (8% modulation depth) and b) sines rotating between the left and right channels. Rotation was simulated through periodic changes in the relative intensity of the left and right channels ($\pm 2.5\text{dB}$) at rates equal to the frequency difference between the two

channels of each two-component stimulus. For the amplitude-modulated stimuli, modulation rates were also equal to the frequency difference between the two channels of the relevant two-component stimuli. The modulation depth (8%) and intensity change ($\pm 2.5\text{dB}$) values were chosen so that the comparison stimuli would sound as similar as possible⁹⁰ to the stimuli of interest.⁹¹ All stimuli were presented dichotically. The purpose of the experiment was to examine:

- a) whether subjects perceived dichotic beats as loudness fluctuations or as stimulus rotation, if at all, and
- b) whether their decision depended on register and/or the frequency difference between left and right channels.

⁹⁰ No formal experiment was conducted to determine the modulation depth and intensity change values that resulted in stimuli similar sounding to the stimuli of interest. The values were chosen based on repeated trials by the experimenter and one additional subject.

⁹¹ Stimuli of interest were the stimuli with different frequencies between the left and right channels (candidates for dichotic beats).

Table 6.11
List of Stimuli Used in Experiment 3, Testing Possible Beats Due to Dichotic Frequency Differences

Underlined stimuli were the stimuli of interest (i.e. candidates for dichotic beats). All other stimuli were included as comparison stimuli. The rotating stimuli were created by changing periodically the intensity relationship between the two channels ($\pm 2.5\text{dB}$).
For the amplitude modulated stimuli, AM-depth = 8%.

Signal group 1: 300Hz	Signal group 2: 1700Hz
Left: 300Hz - Right: 300Hz	Left: 1700Hz - Right: 1700Hz
<u>Left: 300Hz - Right: 301Hz</u> L & R: 300.5Hz - <i>Rotation Rate</i> : 1Hz L & R: 300.5Hz - <i>AM-rate</i> : 1Hz	<u>Left: 1700Hz - Right: 1701Hz</u> L & R: 1700.5Hz - <i>Rotation Rate</i> : 1Hz L & R: 1700.5Hz - <i>AM-rate</i> : 1Hz
<u>Left: 300Hz - Right: 302Hz</u> L & R: 301Hz - <i>Rotation Rate</i> : 2Hz L & R: 301Hz - <i>AM-rate</i> : 2Hz	<u>Left: 1700Hz - Right: 1702Hz</u> L & R: 1701Hz - <i>Rotation Rate</i> : 2Hz L & R: 1701Hz - <i>AM-rate</i> : 2Hz
<u>Left: 300Hz - Right: 303Hz</u> L & R: 301.5Hz - <i>Rotation Rate</i> : 3Hz L & R: 301.5Hz - <i>AM-rate</i> : 3Hz	<u>Left: 1700Hz - Right: 1703Hz</u> L & R: 1701.5Hz - <i>Rotation Rate</i> : 3Hz L & R: 1701.5Hz - <i>AM-rate</i> : 3Hz
<u>Left: 300Hz - Right: 310Hz</u> L & R: 305Hz - <i>Rotation Rate</i> : 10Hz L & R: 305Hz - <i>AM-rate</i> : 10Hz	<u>Left: 1700Hz - Right: 1710Hz</u> L & R: 1705Hz - <i>Rotation Rate</i> : 10Hz L & R: 1705Hz - <i>AM-rate</i> : 10Hz

The ten subjects listened to the stimuli through headphones and were presented with a total of eight categorization tasks in eight consecutive windows like the one appearing in Figure 6.7.

Each window contained three *model stimuli* at the top and four *experiment stimuli* at the bottom. The task was to group all experiment stimuli under the model stimuli. Subjects could listen to all stimuli as often as needed before (and after) making a decision, by right-clicking on each stimulus box. Left-clicking

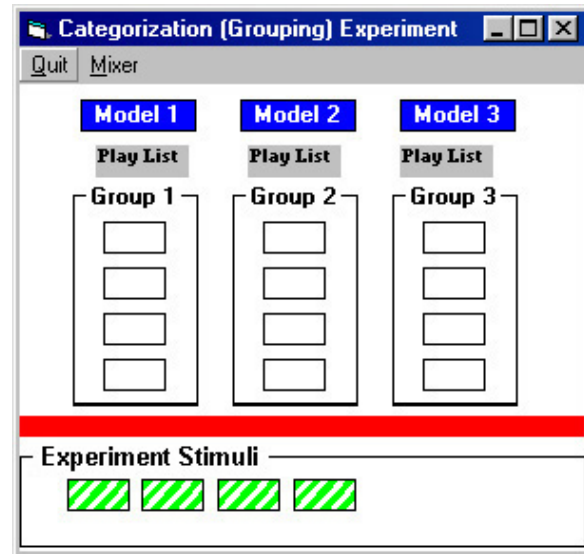


Figure 6.7: Computer interface for the categorization tasks in Experiment 3.

on an experiment stimulus allowed dragging and placing it under the relevant model stimulus. After placing all experiment stimuli in the chosen blank boxes, subjects would click on the *Quit* menu at the top left of the window to record their responses and move to the next categorization task/window. There was a training session to familiarize the subjects with the stimuli and the computer interface.

Table 6.12 displays the *model* and the *experiment stimuli* for the eight, randomly presented, categorization tasks (four for the 300Hz stimuli groups and four for the 1700Hz stimuli groups).

Table 6.12 Experiment and Model Stimuli for the Eight Categorization Tasks in Experiment 3 There were 4 categorization tasks for the 300Hz stimuli groups and 4 for the 1700Hz stimuli groups. The stimuli lists contained the dichotic stimuli of interest (<u>underlined</u>) along with the three stimuli in each respective model list.		
	Experiment Stimuli	Model Stimuli
1	<u>Left: 300Hz - Right: 301Hz</u> Left: 300Hz - Right: 300Hz L & R: 300.5Hz - <i>Rotation Rate</i> : 1Hz L & R: 300.5Hz - <i>AM-rate</i> : 1Hz	Left: 300Hz - Right: 300Hz L & R: 300.5Hz - <i>Rotation Rate</i> : 1Hz L & R: 300.5Hz - <i>AM-rate</i> : 1Hz
2	<u>Left: 300Hz - Right: 302Hz</u> Left: 300Hz - Right: 300Hz L & R: 301Hz - <i>Rotation Rate</i> : 2Hz L & R: 301Hz - <i>AM-rate</i> : 2Hz	Left: 300Hz - Right: 300Hz L & R: 301Hz - <i>Rotation Rate</i> : 2Hz L & R: 301Hz - <i>AM-rate</i> : 2Hz
3	<u>Left: 300Hz - Right: 303Hz</u> Left: 300Hz - Right: 300Hz L & R: 301.5Hz - <i>Rotation Rate</i> : 3Hz L & R: 301.5Hz - <i>AM-rate</i> : 3Hz	Left: 300Hz - Right: 300Hz L & R: 301.5Hz - <i>Rotation Rate</i> : 3Hz L & R: 301.5Hz - <i>AM-rate</i> : 3Hz
4	<u>Left: 300Hz - Right: 310Hz</u> Left: 300Hz - Right: 300Hz L & R: 305Hz - <i>Rotation Rate</i> : 10Hz L & R: 305Hz - <i>AM-rate</i> : 10Hz	Left: 300Hz - Right: 300Hz L & R: 305Hz - <i>Rotation Rate</i> : 10Hz L & R: 305Hz - <i>AM-rate</i> : 10Hz
5	<u>Left: 1700Hz - Right: 1701Hz</u> Left: 1700Hz - Right: 1700Hz L & R: 1700.5Hz - <i>Rotation Rate</i> : 1Hz L & R: 1700.5Hz - <i>AM-rate</i> : 1Hz	Left: 1700Hz - Right: 1700Hz L & R: 1700.5Hz - <i>Rotation Rate</i> : 1Hz L & R: 1700.5Hz - <i>AM-rate</i> : 1Hz
6	<u>Left: 1700Hz - Right: 1702Hz</u> Left: 1700Hz - Right: 1700Hz L & R: 1701Hz - <i>Rotation Rate</i> : 2Hz L & R: 1701Hz - <i>AM-rate</i> : 2Hz	Left: 1700Hz - Right: 1700Hz L & R: 1701Hz - <i>Rotation Rate</i> : 2Hz L & R: 1701Hz - <i>AM-rate</i> : 2Hz
7	<u>Left: 1700Hz - Right: 1703Hz</u> Left: 1700Hz - Right: 1700Hz L & R: 1701.5Hz - <i>Rotation Rate</i> : 3Hz L & R: 1701.5Hz - <i>AM-rate</i> : 3Hz	Left: 1700Hz - Right: 1700Hz L & R: 1701.5Hz - <i>Rotation Rate</i> : 3Hz L & R: 1701.5Hz - <i>AM-rate</i> : 3Hz
8	<u>Left: 1700Hz - Right: 1710Hz</u> Left: 1700Hz - Right: 1700Hz L & R: 1705Hz - <i>Rotation Rate</i> : 10Hz L & R: 1705Hz - <i>AM-rate</i> : 10Hz	Left: 1700Hz - Right: 1700Hz L & R: 1705Hz - <i>Rotation Rate</i> : 10Hz L & R: 1705Hz - <i>AM-rate</i> : 10Hz

The *experiment stimuli* lists contained the dichotic stimuli of interest (underlined) along with the three stimuli in the respective *model stimuli* lists. These three stimuli were included to ensure that subjects could recognize identities and differences among the model stimuli, demonstrating that they had understood the task and that their decisions regarding the stimuli of interest were based on the desired criteria.

Results

Three data points were obtained by each subject for each experiment stimulus in each categorization task. A *1* was assigned to the model stimulus the experiment stimulus was matched to and a *0* was assigned to each model stimulus left unselected. These data points were added across subjects giving three values for each experiment stimulus indicating how many times it was categorized under each model stimulus. Therefore, there was one multinomial dependent variable with 3 levels (3 model stimuli).

The Linear Multinomial ANOVA statistical model was used for the analysis (Woodward et al., 1990: 431-467) with 3 within-subjects factors:

- a) *Stimulus Category* (fixed factor with 4 levels: Steady Sine - Left-to-Right Rotating Sine - AM Sine - Dichotic Tone)

- b) *Rotation Rate/AM Rate/Frequency Difference* (random factor nested under *Stimulus Category* with 4 levels: 1Hz - 2Hz - 3Hz - 10Hz) and
- c) *Frequency Range* (random factor nested under *Rotation Rate* with 2 levels: 300Hz and 1700Hz, representing the Under 500Hz / Over 1500Hz ranges respectively).

Data analysis was performed using GANOVA-4, a custom statistical software incorporating the Linear Multinomial ANOVA model, part of Woodward et al. (1990). The software interpreted the design as having four multinomial dependent variables: the three within-subject factors listed above and the one dependent variable nested under the last factor.

Therefore, the four multinomial dependent variables were:

- a) *Stimulus Category*, 4 levels,
- b) *Rotation rate/AM Rate/Frequency Difference*, 4 levels, nested under *Stimulus Category*,
- c) *Frequency Range*, 2 levels, nested under *Rotation Rate* and
- d) *Model Stimulus*, 3 levels, nested under *Frequency Range*.

Separate analyses were performed on the data for each of the four levels in the *Stimulus Category* variable, with $\alpha = 0.0125$ each for a total experiment-wise error $\alpha_{ew} = 0.05$. The reason for the separate analyses was that the *Stimulus Category* variable

was not orthogonal to the other dependent variables. The orthogonality assumption was satisfied for the remaining three dependent variables. The Pearson significance test was used, comparing the Pearson test statistic to the Chi-square distribution for the degrees of freedom appropriate to each hypothesis testing.⁹² Due to the very low number of subjects, the inferential power of the analysis was severely compromised. The reported results will have to be confirmed by a future study using ten times the number of subjects for the power to exceed $P = 0.8$.

Figures 6.8 to 6.10 display the categorization results for the stimuli that functioned as both model and experiment stimuli [Steady Sines (ST), Left-to-Right Rotating Sines (LR), and AM Sines (AM)] for the four rotation/modulation rates and the <500Hz (300Hz) and >1500Hz (1700Hz) stimuli groups.

⁹² The Pearson test and test statistic for the linear multinomial ANOVA model are defined in Woodward et al. (1990: 437).

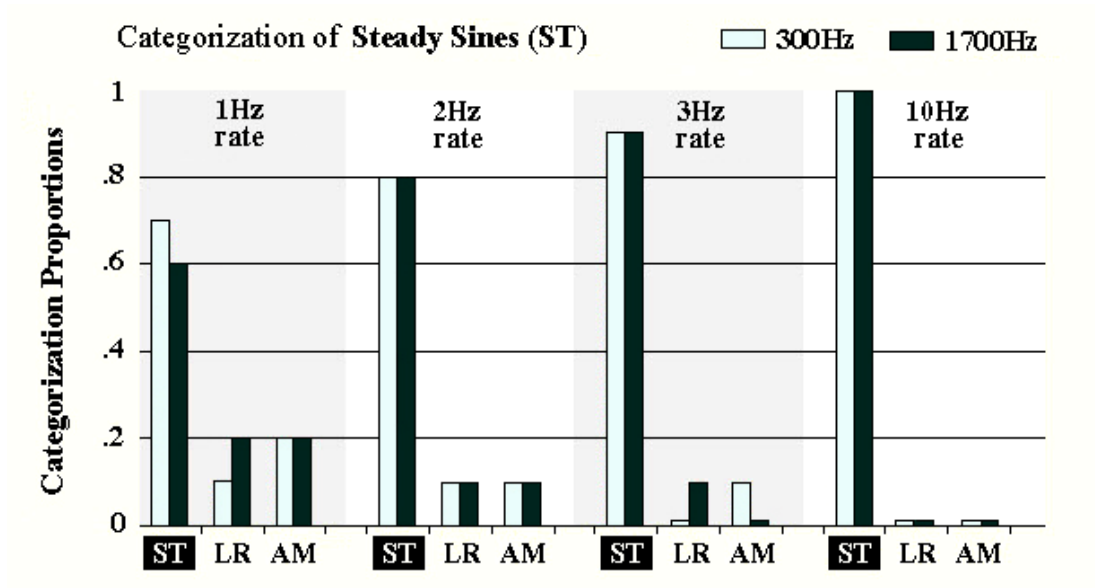


Figure 6.8: Categorization results for the 300/1700Hz steady sines (ST) in Experiment 3. The stimuli were compared to a) themselves (ST), b) 300/1700Hz left-to-right rotating sines (LR), and c) 300/1700Hz amplitude-modulated sines (AM), at four different rotation/modulation rates (AM-depth = 8%).

Figure 6.8 displays the categorization results for the 300/1700 steady sines (ST). There was no significant difference between the two frequency groups (*Pearson test statistic* = 0.015; *df* = 1; $p > 0.500$), indicating that the categorization response patterns for the steady sines were the same for both, the 300Hz and the 1700Hz centered stimuli. Subjects were clearly able to correctly identify the 300/1700Hz steady sine (ST) stimulus among 300/1700Hz rotating and amplitude-modulated stimuli for all 4 rotation / modulation rates (*Pearson test statistic* = 1.597; *df* = 3; $p > 0.500$).

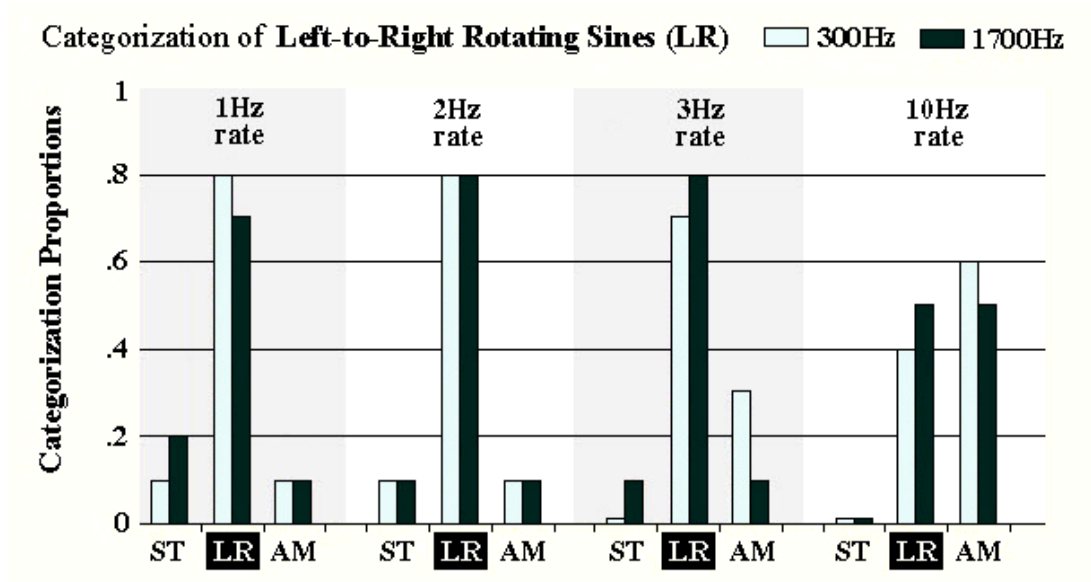


Figure 6.9: Categorization results for the 300/1700Hz left-to-right rotating sines (LR) in Experiment 3. The stimuli were compared to a) themselves (LR), b) 300/1700Hz steady sines (ST), and c) 300/1700Hz amplitude-modulated sines (AM), at four different rotation/modulation rates (AM-depth = 8%).

Figure 6.9 displays the categorization results for the 300/1700 rotating sines (LR). There was again no significant difference between the two frequency groups (*Pearson test statistic* = 0.018; *df* = 1; *p* > 0.500), indicating that the categorization response patterns for the left-to-right rotating sines were the same for both, the 300Hz and the 1700Hz centered stimuli. This was also the case for rotation / modulation rates of 1-3Hz (*Pearson test statistic* = 0.043; *df* = 2; *p* > 0.500), indicating that subjects were able to correctly identify the 300/1700Hz left-to-right rotating sine

(LR) among steady and amplitude-modulated 300/1700Hz sines for rotation / modulation rates $\leq 3\text{Hz}$. For the 10Hz rotation / modulation rate, however, the rotating stimulus was categorized differently (*Pearson test statistic* = 7.128; *df* = 1; *p* = 0.008) and, as indicated in Figure 6.9, it was confused with the amplitude-modulated stimulus. Similar results were obtained from the categorization of the amplitude-modulated stimuli displayed in Figure 6.10, below.

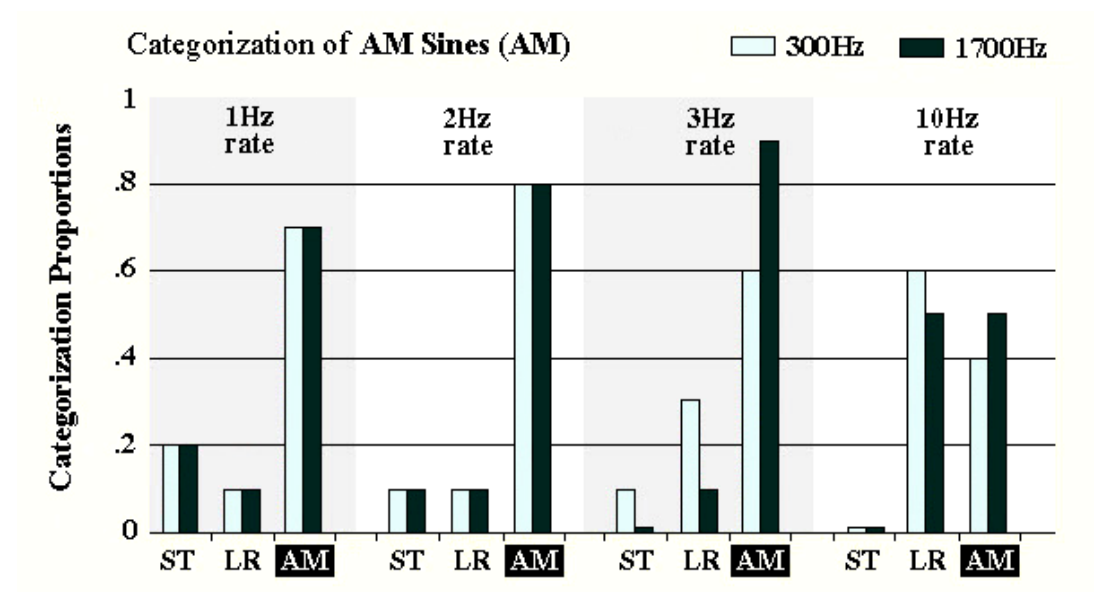


Figure 6.10: Categorization results for the 300/1700Hz amplitude-modulated sines (AM) in Experiment 3. The stimuli were compared to a) themselves (AM), b) 300/1700Hz steady sines (ST), and c) 300/1700Hz left-to-right rotating sines (LR), at four different rotation/modulation rates (AM-depth = 8%).

Figure 6.10 displays the categorization results for the 300/1700 amplitude-modulated sines (AM). There was no significant difference between the two frequency groups (*Pearson test statistic* = 0.296; *df* = 1; *p* > 0.500). As was the case

with the left-to-right rotating stimuli, subjects were able to correctly identify the 300/1700Hz AM sine stimulus among steady and rotating 300/1700Hz sines *only* for rotation / modulation rates $\approx \leq 3\text{Hz}$ (*Pearson test statistic* = 0.133; *df* = 2; *p* > 0.500). For the 10Hz rotation / modulation rate, the amplitude-modulated stimulus was categorized differently (*Pearson test statistic* = 6.750; *df* = 1; *p* = 0.009) and was confused with the rotating stimulus.

To summarize, Figures 6.8, 6.9, & 6.10 indicate that subjects were able to identify the subtle differences between the model stimuli for both frequency ranges (<500Hz and >1700Hz, represented by the frequency values of 300Hz and 1700Hz respectively) and for all rotation / modulation rates with one important exemption: When the rotation / modulation rate increased beyond 3Hz (i.e. 10Hz), subjects confused the rotating stimuli with the amplitude-modulated stimuli and vice versa, for both frequency ranges. This observed confusion is consistent with earlier observations (Lane, 1925) and supports the hypothesis that the beating sensation reported in previous dichotic experiments may have been a misidentified rotating sensation. Figure 6.11 displays the categorization results for the primary stimuli of interest (dichotic stimuli with different frequencies in the left and right channels) providing further support to this hypothesis.

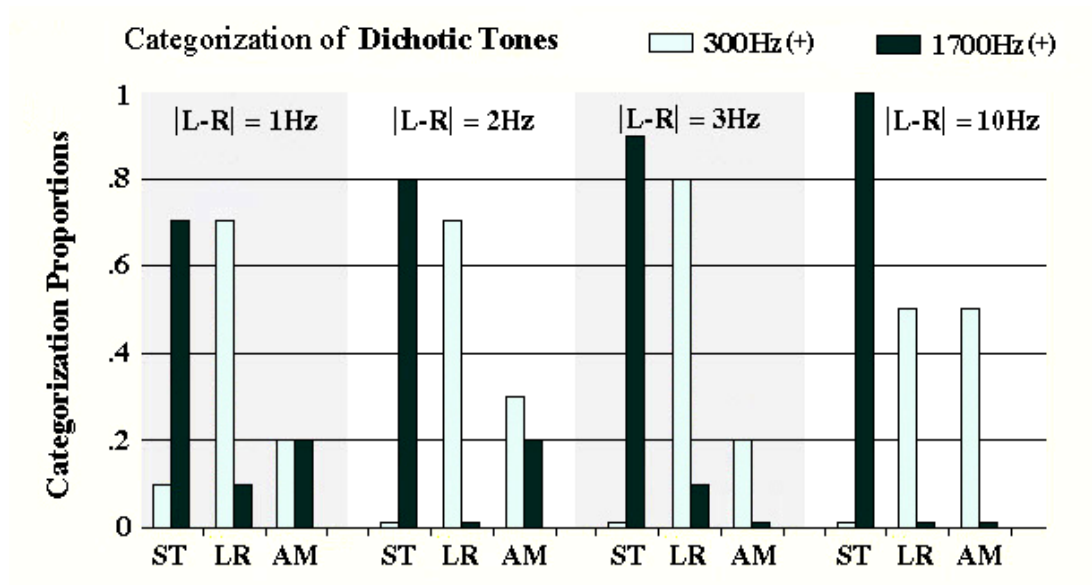


Figure 6.11: Categorization results for the two-component dichotic tones in Experiment 3. Left channel = 300/1700Hz - Right channel = 300/1700Hz + 1Hz, 2Hz, 3Hz, or 10Hz. The dichotic tones were the primary stimuli of interest and were grouped by the subjects under one of the three model-stimuli categories: a) steady sines (ST), b) left-to-right rotating sines (LR), and c) amplitude-modulated sines (AM), for four different rotation/modulation rates (AM-depth = 8%).

Figure 6.11 illustrates a difference in the categorization patterns of the dichotic tones for the two frequency groups (300Hz - 1700Hz). Analysis confirmed a clear frequency effect (*Pearson test statistic* = 31.115; $df = 1$; $p < 0.001$) indicating that subjects categorized the dichotic stimuli differently for the 300Hz and 1700Hz ranges.

In the 300Hz range (light bars), subjects grouped the dichotic tones with frequency differences (rotation / modulation rates) ≤ 3 Hz under the rotating model-stimuli (*Pearson test statistic* = 0.091; $df = 2$; $p > 0.500$). The dichotic tone with 10Hz frequency difference was grouped differently (*Pearson test statistic* = 6.346; $df = 1$; p

= 0.012), with the responses being shared between the rotating and the amplitude-modulated model-stimuli. There is a strong similarity between these results and the categorization results for the left-to-right rotating tones. It indicates that, in the 300Hz range, two-component dichotic stimuli with different frequencies in each ear are treated perceptually as rotating sines.

In the 1700Hz range (dark bars), subjects grouped the dichotic tones under the steady sine model-stimuli for all frequency differences (rotation / modulation rates) (*Pearson test statistic* = 0.598; *df* = 3; *p* > 0.500). There is a strong similarity between these results and the categorization results for steady tones. It indicates that, in the 1700Hz range, two-component dichotic stimuli with different frequencies in each ear are treated perceptually as steady sines, with the frequency difference between ears introducing no perceptual difference.

The results of Experiments 2a, 2b, 2c, & 3 support the hypothesis against the claim of sound interference products arising due to wave interaction at a neural level. They indicate that the beating sensation reported in previous dichotic experiments must have simply been a misidentified rotating sensation, a manifestation of the phase-based sound localization mechanism discussed in the literature (i.e. Blauert, 1996; Hartmann, 1999).

6.2.5 Summary

The energy content of amplitude fluctuations has been questioned by previous experiments where sound interference products appear to arise at a neural level. The possibility of interference products arising from neural wave interaction was examined through a series of dichotic experiments. It was shown that the perception of sound interference in terms of loudness changes, beating, or roughness does not arise when sound waves are not permitted to physically interfere. Three sets of dichotic experiments confirmed this claim.

The first experiment examined the possibility of beating when a sine stimulus is presented to both ears through headphones. In this case, the phenomenon of diplacusis causes a pitch difference between the two ears that would introduce beating if neural wave interaction could give rise to interference products. However, no such beating was observed.

The second set of experiments examined the effect of phase on dichotically presented sine stimuli. Results indicated that changes in the phase relationship between the left and right channels of dichotic sine signals give rise to perceptions that are incompatible with the interference principle and consistent with studies that have

established phase differences between ears as sound localization cues. It was shown that:

- Changing the phase relationship between the left and right channels of a sinusoidal stimulus presented dichotically does not change its loudness (Experiment 2a).
- For frequencies <500Hz (200Hz & 300Hz), dichotic phase changes affect the perceived direction of a stimulus (Experiment 2b), with 180° phase difference resulting in a wider stereo image (Experiment 2c).
- For frequencies >1500Hz (1700Hz & 2500Hz), dichotic phase changes introduce no perceptual change (Experiments 2b & 2c).

The results support the claim that the beating-like sensation reported in previous dichotic experiments involving two sines with slightly different frequencies is not an indication of interference products arising at a neural level but the manifestation of a sound localization mechanism based on phase difference cues.

The last experiment (Experiment 3) was designed to examine this claim. It was shown that, in dichotic explorations of beats, the constantly shifting phase between the sine components presented in the two ears results in a constantly shifting localization of the two-component complex stimulus. For low frequency differences ($\approx \leq 3\text{Hz}$) this shift is interpreted as a rotation of the sound inside the subject's head and, for higher frequency differences ($\approx \geq 3\text{Hz}$) it is interpreted as a timbral

fluctuation that is often confused by the subjects with loudness fluctuation / beating. This confusion was shown to always arise when comparing rotating and modulated sines with rotation / modulation rates $\approx \geq 3\text{Hz}$. Consistent with sound localization studies and contrary to the interference principle, it was shown that two-component dichotic stimuli with different frequencies in each ear are treated perceptually as rotating sines for frequencies $< 500\text{Hz}$ (300Hz).⁹³ For frequencies $> 1500\text{Hz}$ (1700Hz) they are treated as steady sines, with the frequency difference between ears introducing no perceptual difference.

The findings support the hypothesis against sound interference products arising due to neural wave interaction. They are consistent with the arguments for the energy content of amplitude fluctuation and indicate that the beating sensation reported in previous dichotic experiments must have been a misidentified rotating sensation.

⁹³ Perrott & Musicant's (1977) results support a quite different interpretation of the relationship between rotating tones and dichotic (or binaural) beats. Their experimental results could not be replicated by the present study so the proposed interpretation does not take into account Perrott & Musicant's report.

6.3 Energy Content of Amplitude Fluctuations – Alternative Signal Representation

6.3.1 Theoretical Value of the Amplitude Fluctuation Energy in Two-Component Complex Signals

The experiment results in Section 6.2 indicate that sound interference products do not arise when sound waves are not allowed to physically interact. Physical interaction of sound waves gives rise to amplitude fluctuations. The perception of sound interference products may therefore be a manifestation of amplitude fluctuation energy. Four main points have been outlined in Sections 5.2 & 5.4, arguing for the wave nature of amplitude fluctuations (a special case of wave modulations) and their energy content:

- Amplitude fluctuations can transfer energy to systems (i.e. machinery, resonators, etc.) whose natural frequency matches the fluctuation rate.
- Amplitude fluctuations of acoustic waves correspond to changes in the vibration velocity of a mass at a rate equal to the rate of fluctuation, indicating consumption / radiation of energy at the same rate.

- Amplitude fluctuations entering a dispersive medium separate from the wave that carries them and travel independently at an observable rate and strength.
- Successive portions of a diffracted wave-front are determined by the envelope of secondary wavelets, indicating that amplitude fluctuations and the wave envelopes that describe them behave as waves.

The above points suggest that amplitude fluctuations resulting from the superposition of two sine waves (f_1, A_1, ϕ_1 , and f_2, A_2, ϕ_2) are waves with parameters $f_{\text{mod}}, A_{\text{mod}}, \phi_{\text{mod}}$ that are related to the parameters of the interfering sines. Amplitude fluctuations are a special case of wave modulations and carry energy at the fluctuation rate $f_{AF} = 2f_{\text{mod}} = |f_2 - f_1|$ (difference frequency). Hamilton (1998) provides the generalized expressions for the pressure at the difference frequency in the case of directional (1998: 222), Gaussian (1998: 223-224), and piston sources (1998: 225). The general form of the quasilinear solution for the difference frequency obtained by these expressions requires numerical integration. The use of sophisticated computational techniques is necessary because of the highly oscillatory nature of the integrands (Hamilton, 1998: 225).

As a simplified special case we can assume a single point source oscillating at two frequencies f_1, f_2 , with the same amplitudes $A_1 = A_2 = A$, and zero initial

phases $\phi_1 = \phi_2 = 0$ rads. Eq. (6.1) describes the two-component wave originating from the point source as a single frequency with slowly varying amplitude.

$$y(t) = A \cos 2\pi f_1 t + A \cos 2\pi f_2 t = 2A \cos 2\pi f_{\text{mod}} t (\cos 2\pi f_{\text{av}} t) \quad \text{Eq. (6.1)}$$

$$\text{where } \pi = 3.14\dots, \quad f_{\text{av}} = \frac{f_1 + f_2}{2}, \quad \text{and} \quad f_{\text{mod}} = \frac{|f_2 - f_1|}{2}.$$

Using the identity $\omega = 2\pi f$ we can rewrite Eq. (6.1) as an expression of angular frequency ω to facilitate the power calculations that follow:

$$y(t) = A \cos(\omega_1 t) + A \cos(\omega_2 t) = 2A \cos \omega_{\text{mod}} t (\cos \omega_{\text{av}} t) \quad \text{Eq. (6.2)}$$

$$\text{where } \omega_{\text{av}} = \frac{\omega_2 + \omega_1}{2}, \quad \text{and} \quad \omega_{\text{mod}} = \frac{|\omega_2 - \omega_1|}{2}.$$

In general, the power of a complex wave is not equal to the sum of the powers of its sine components. As already mentioned in Section 5.1, the linearity of superposition does not apply to energy quantities of wave motion, which are quadratic with respect to field variables such as vibration velocity. However, in the case of interference (i.e. Eq. 6.2), if the response time of the receiving system (i.e. air molecules, eardrum, etc.) is long relative to the period of the complex wave's amplitude fluctuations, the power of the superposition is equal to the sum of the powers of the individual components (Towne, 1967: 212). In this case, a theoretical

value for the energy content of amplitude fluctuations can be calculated by applying the energy conservation law⁹⁴ to the two sides of Eq. (6.2).

The power P of a point source pulsating with angular frequency ω and amplitude A inside a non-dispersive medium (i.e. air) is:

$$P = \frac{\omega^2 \rho A^2}{8\pi c} \quad \text{Eq. (6.3)}^{95}$$

where ρ is the equilibrium density of air and c is the speed of sound in air.

Based on Eq. (6.3), the power P_L of the left side of Eq. (6.2) is:

$$P_L = \frac{\omega_1^2 \rho A^2 + \omega_2^2 \rho A^2}{8\pi c} \quad \text{Eq. (6.4)}$$

while the power P_R of the right side of Eq. (6.2) is:

$$P_R = \frac{\omega_{av}^2 \rho (2A)^2}{8\pi c} \cos^2 \omega_{mod} t = \left(\frac{\omega_{av}^2 \rho (2A)^2}{8\pi c} \right) \frac{1}{2} (1 + \cos 2\omega_{mod} t) =$$

⁹⁴ The energy conservation law states that, in the absence of external forces, the sum of kinetic and potential energies in a vibration remains constant.

$$\frac{\omega_{av}^2 \rho A^2}{4\pi c} + \frac{\omega_{av}^2 \rho A^2}{4\pi c} \cos 2\omega_{mod} t = \frac{\omega_{av}^2 \rho A^2}{4\pi c} = \frac{\left(\frac{\omega_1 + \omega_2}{2}\right)^2 \rho A^2}{4\pi c} = \frac{\omega_1^2 \rho A^2 + \omega_2^2 \rho A^2 + 2\omega_1 \omega_2 \rho A^2}{16\pi c}.$$

$$\text{So } P_R = \frac{\omega_1^2 \rho A^2 + \omega_2^2 \rho A^2 + 2\omega_1 \omega_2 \rho A^2}{16\pi c}. \quad \text{Eq. (6.5)}$$

Eqs. (6.4) & (6.5) show that $P_L > P_R$. The present study interprets the power difference $P_L - P_R$ as the power radiated by the amplitude fluctuations P_{AF} .

$$P_{AF} = P_L - P_R = \frac{\omega_1^2 \rho A^2 + \omega_2^2 \rho A^2}{8\pi c} - \frac{\omega_1^2 \rho A^2 + \omega_2^2 \rho A^2 + 2\omega_1 \omega_2 \rho A^2}{16\pi c} = \frac{\omega_1^2 \rho A^2 + \omega_2^2 \rho A^2 - 2\omega_1 \omega_2 \rho A^2}{16\pi c} = \frac{(\omega_1 - \omega_2)^2 \rho A^2}{16\pi c}$$

$$\text{So } P_{AF} = \frac{\omega_{AF}^2 \rho A^2}{16\pi c} = \frac{\omega_{AF}^2 \rho (A\sqrt{0.5})^2}{8\pi c} = \frac{\omega_{AF}^2 \rho A_{AF}^2}{8\pi c} \quad \text{Eq. (6.6)}$$

According to Eq. (6.3), Eq (6.6) describes the power radiated by a point source pulsating with angular frequency $\omega_{AF} = |\omega_1 - \omega_2|$ (rate of amplitude fluctuation) and amplitude $A_{AF} = \sqrt{0.5}A$, where ω_1 , ω_2 , and A represent the angular frequencies and common amplitude of the interfering sines respectively. Eq. (6.7) expresses the power

⁹⁵ Fletcher & Rossing, 1998: 172.

of the amplitude fluctuations as a function of the sum of the powers radiated by the two sines independently (P_L):

$$P_{AF} = P_L \frac{(\omega_1 - \omega_2)^2}{2(\omega_1^2 + \omega_2^2)} \quad \text{Eq. (6.7)}$$

The more general case of sines with arbitrary amplitude relationship can be examined based on Eq. (6.8):

$$y(t) = A_1 \cos \omega_1 t + A_2 \cos(\omega_2 t) = \sqrt{(A_1^2 + A_2^2 + 2A_1 A_2 \cos \omega_{AF} t)} \cos(\omega_1 t + \phi(t)) \quad \text{Eq. (6.8)}^{96}$$

where $\phi(t)$ is a time-dependent phase with:

$$\tan \phi(t) = \frac{A_2 \sin \omega_{AF} t}{A_1 + A_2 \cos \omega_{AF} t}. \quad \text{Eq. (6.9)}$$

The present study will be limited to the examination of the case described by Eqs. (6.2) & (6.6).

⁹⁶ Derived from Fletcher & Rossing (1998: 9) by setting: $\phi_1 = \phi_2 = 0$.

6.3.2 Alternative Signal Representation Accounting for the Energy Content of Amplitude Fluctuations

Two-dimensional complex signals (time included) do not account for energy at the rate of amplitude fluctuation. As already indicated,⁹⁷ such signals neither account for the negative amplitude values predicted by the mathematical expression of amplitude fluctuation (Eq. (6.2)) nor offer a consistent measure of the energy content of sine waves across frequencies. An alternative representation is proposed, based on the rotating-vector method (Crawford, 1973: 308). It includes both the sine (imaginary) and cosine (real) terms describing a vibration, and results in spiral sine signals and twisted-spiral complex signals. This representation alleviates the negative-amplitude and consistent-energy-representation problems while mapping the twisting parameters onto the amplitude fluctuation ones.

The general equation used to describe a simple harmonic motion is:

$$y(t) = A \cos(\omega t) \quad \text{Eq. (6.10)}$$

⁹⁷ Section 5.2.1.

Eq. (6.10) is the real part of the complex (i.e. including imaginary numbers) equation of motion (e : natural logarithmic base = 2.71828...; $i = \sqrt{-1}$):

$$y_c(t) = Ae^{i\omega t} \quad \text{Eq. (6.11)}$$

Figure 6.12 illustrates Eq. (6.11) as a vector $e^{i\omega t}$ with length A and angular frequency ω , rotating counter-clockwise on the complex plane. One of the advantages of such a representation is that the area scanned by the vectors is proportional to frequency. As opposed, therefore, to the traditional two-dimensional representation (Figure 5.1), the signal resulting from graphing Eq. (6.11) represents correctly the energy content of a wave.

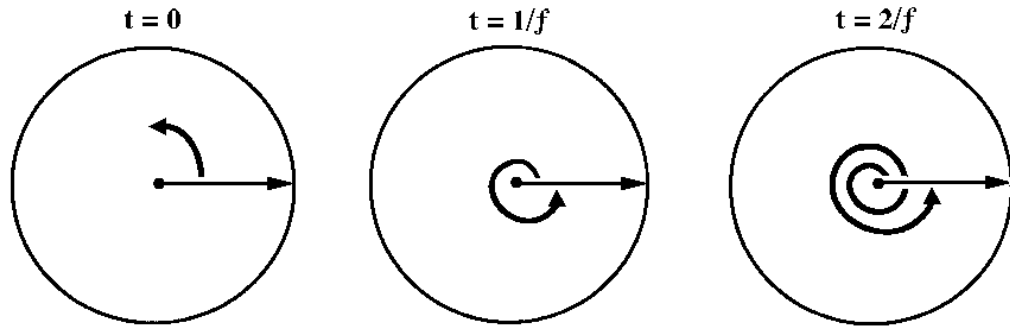


Figure 6.12: Graph of a vector $e^{i\omega t}$ with length A , rotating on the complex plane with angular frequency ω (Eq. (6.11)). The vector goes through its initial position (left) $f=\omega/2\pi$ times each second. The snapshots (middle & right) are taken at times $t=1/f=T$ and $t=2/f=2T$ respectively. In the rotating vector representation, the area scanned by the vector is proportional to frequency.

In Figure 6.12, the time axis is perpendicular to the plane of the figure, passing through the center of the circle. Projecting the motion of the rotating vector on a plane containing the time axis (real plane) represents the motion described by Eq. (6.10) and results in the familiar sinusoidal signal (Figure 5.1).

Representing the motion described by Eq. (6.11) in three dimensions results in spiral sine signals and twisted-spiral complex signals. Two examples of the twisted spiral representation are illustrated in Figure 6.13.

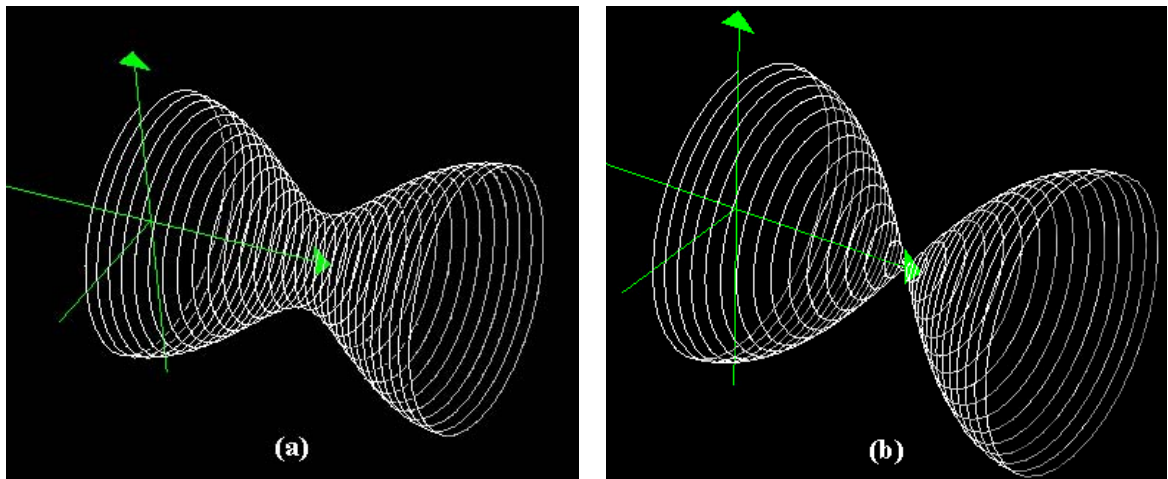


Figure 6.13: Twisted spiral representation of a two-component complex signal with 1Hz frequency difference between components and (a) unequal or (b) equal amplitudes. The horizontal arrow represents time, the vertical arrow represents the sine function (imaginary axis), and the third axis represents the cosine function (real axis).

Figure 6.13 illustrates a two component complex signal with 1Hz frequency difference between components and (a) unequal or (b) equal amplitudes, as a twisted spiral.

Figure 6.14 (below) is an alternative illustration of the *twisted signals* hypothesis for the case of a two-component complex signal with $f_1 = 10\text{Hz}$; $f_2 = 12\text{Hz}$; $A_1 = A_2$; ϕ_1, ϕ_2 .

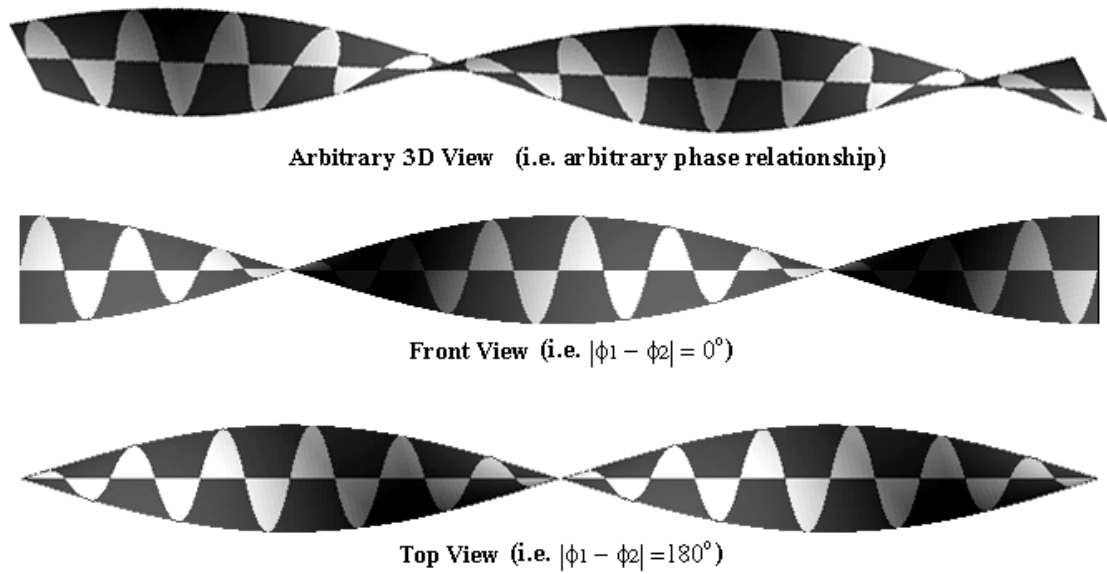


Figure 6.14: Alternative illustration of the twisted-signal hypothesis.

In Figure 6.14:

- Modulation rate f_{mod} is represented by the number of twists around the horizontal axis: $f_{\text{mod}} = (f_2 - f_1)/2 = 1\text{Hz}$ (1 twist).
- Degree of amplitude fluctuation is represented by the twist radius.

- Amplitude fluctuation rate f_{af} is represented by the number of fluctuations formed when the twisted signal is projected on two dimensions:

$$f_{af} = 2f_{\text{mod}} = 2\text{Hz} \text{ (2 fluctuations).}$$

Twisted three-dimensional signals of two-component complex waves represent more accurately all wave parameters, including the parameters of amplitude fluctuation. Furthermore, twisted signals paint a more realistic picture of wave propagation. Visualization of standing waves in air columns has revealed swirling or twisting air flows (i.e. Campbell & Greated, 1987: 288). A series of examples of three-dimensional twisted spiral complex signals are displayed in the Appendix, illustrating the usefulness of the proposed representation.

If the radiated power P'_{AF} of the amplitude fluctuations is calculated based on the expression: $A_{\text{mod}} = 2A \cos \omega_{\text{mod}}$ (which treats modulations / beats as independent waves) rather than on the energy conservation law, we get:

$$P'_{AF} = \frac{\omega_{\text{mod}}^2 \rho (2A)^2}{8\pi c} = \frac{\left(\frac{\omega_1 - \omega_2}{2}\right)^2 \rho (2A)^2}{8\pi c} = \frac{(\omega_1 - \omega_2)^2 \rho A^2}{8\pi c} = \frac{\omega_{AF}^2 \rho A^2}{8\pi c} \quad \text{Eq. (6.12)}$$

Eqs. (6.6) & (6.12) indicate that: $P_{AF} = P'_{AF} / 2$. It is possible to explain this discrepancy using the concept of twisted complex signals. The energy required to gradually *twist* the 1-second-long signal (Figure 6.14) $\omega_{\text{mod}} / 2\pi$ times along the x axis is equal to half the energy required to *turn* the same signal $\omega_{\text{mod}} / 2\pi$ times along the x axis. Eq. (6.12) represents the power related to the turn, rather than the twist of the 1-second-long signal. This may explain why the power calculated on the basis of the frequency and amplitude values of A_{mod} (Eq. (6.12)) is two times larger than the value calculated on the basis of the energy conservation law. Eq. (6.6) represents the proposed theoretical estimate of the power radiated by amplitude fluctuations and is addressed experimentally in the following Section.

6.3.3 Experiment 4: Testing the Theoretical Value of the Amplitude

Fluctuation Energy in Two-Component Complex Signals

Hypothesis

Amplitude fluctuations resulting from the superposition of two sine waves with parameters f_1, A_1, ϕ_1 , and f_2, A_2, ϕ_2 ($A_1 = A_2 = A$; $\phi_1 = \phi_2 = 0$) are waves with frequency: $f_{AF} = |f_2 - f_1|$ and amplitude $A_{AF} = \sqrt{0.5}A$.

Methods

Spectral analysis of the complex signal in Eq. (6.2) using the Fast Fourier Transform (FFT) will reveal two sine components⁹⁸ with frequencies f_1, f_2 and amplitudes $A_1 = A_2 = A$. It will not reveal the predicted amplitude fluctuation component ($f_{AF} = |f_2 - f_1|$; $A_{AF} = \sqrt{0.5}A$), corresponding to the power in Eq. (6.6). FFT analyzes any two-dimensional curve into a family of sine curves. The envelope of a two-dimensional complex signal⁹⁹ is a boundary curve that encloses the area

⁹⁸ If the analysis-bandwidth is $< |f_1 - f_2|$.

⁹⁹ Which can be interpreted as the signal of the amplitude fluctuations.

outlined by all maxima of the motion represented by the two-dimensional signal and includes points that do not belong to the signal. Therefore, the envelope is not a member of the family of sine curves forming the complex two-dimensional signal (Davis & Samuels, 1996: 10) and will consequently not be detected through FFT analysis.

Real-time signal analyzers using parallel analogue or digital filter-banks are more appropriate for the task of identifying a resonance at the amplitude fluctuation rate. A mechanical analogue would be a set of finely tuned resonators. Using three tuning forks (or any other resonators) tuned at the frequencies of the two interfering sines and at the rate of amplitude fluctuation respectively, would simulate the Rücker & Edser experiment (1895, in Beyer, 1999: 149-150) with an additional advantage. Vibration measurements on all three tuning forks would measure the relationship between the amplitudes of the interfering sines and the amplitude of the difference frequency. The most appropriate parameter to measure would be vibration velocity since it is proportional to sound pressure level. However, mounting a vibration-measurement transducer on the resonators would significantly alter their resonant characteristics. A device that makes no contact (i.e. a laser vibrometer) would be more appropriate. Otherwise, the frequencies of the signals to be measured would have to be equal to the resonant frequencies of the respective *resonator-transducer-transducer mounting* systems. Several vibration-measurement transducers are listed in Harris (1998) and in Möser (1998).

Alternatively, if appropriate resonators and transducers are not available, FFT analysis can be combined with amplitude demodulation techniques to extract the envelope of a complex signal.¹⁰⁰ The power spectrum of the envelope can then be calculated, indicating the energy at the amplitude fluctuation rate and any of its possible harmonics. The present study adopts the *Hilbert Transform demodulation* technique.

Hilbert transform demodulation

Eq. (6.13) describes a real signal $y(t)$ with frequency f and slowly varying amplitude $A(t)$. The amplitude envelope $A(t)$ can be extracted from the signal using Eq. (6.14). This equation is a function that forms a complex (i.e. including imaginary terms) analytic (i.e. with a spectrum that has only positive frequency components) signal $z(t)$ corresponding to the real signal $y(t)$.

$$y(t) = A(t)\cos(2\pi ft) \quad \text{Eq. (6.13)}$$

$$z(t) = y(t) + iH[y(t)] \quad \text{Eq. (6.14)}$$

¹⁰⁰ This approach is used to monitor the condition of machinery (Krishnappa, 1998; Courrech, 1998).

where $i = \sqrt{-1}$ and $H[y(t)]$ is the Hilbert transform of $y(t)$:

$$H[y(t)] = A(t) \sin(2\pi ft) \quad \text{Eq. (6.15)}$$

The Hilbert Transform (Eq. (6.15)) provides a 90° phase shift to its input and can be thought of as the imaginary component of the real signal $y(t)$. Complex analytic signals are three-dimensional, similar to the three-dimensional twisted signals proposed by the present study (Figures 4.1, 6.13, & 6.14 and the Appendix). Figure 6.15 displays an example of a complex analytic signal.

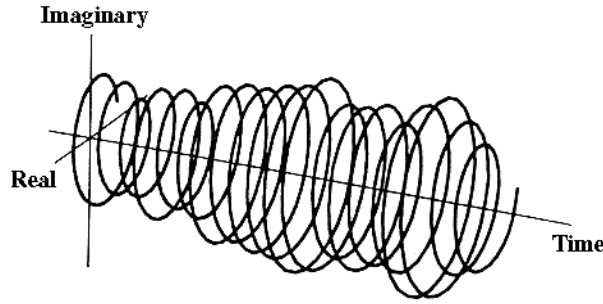


Figure 6.15: Complex analytic signal derived from vibrational measurements on a machinery gear (after Krishnappa, 1998: 709).

Complex analytic signals are useful to the examination of wave propagation for numerous reasons. One of them is that the 90° phase difference between a wave signal and its Hilbert transform reflects the difference between potential and kinetic energy in a propagating wave. In Eq. (6.14), for example, $y(t)$ can be seen as

representing scaled pressure and $H[y(t)]$ as representing scaled velocity.¹⁰¹ The envelope of the signal described by Eq. (6.13) can therefore be seen as tracing the changes in the total (potential + kinetic) instantaneous energy of the signal over time and can be described by the absolute value of the analytic signal in Eq. (6.14):

$$A(t) = \sqrt{y(t)^2 + H[y(t)]^2} \quad \text{Eq. (6.16)}^{102}$$

In conclusion, the envelope of a signal such as $y(t) = A(t)\cos(2\pi ft)$ can be extracted by computing the signal's Hilbert transform and performing the operation in Eq. (6.16).

Figure 6.16 is a block diagram of the envelope extraction process,¹⁰³ including signal generation and windowed frequency zoom analysis.

¹⁰¹ The scales are the square root of compressibility and the square root of density of the propagation medium respectively.

¹⁰² Feldman, 1994: 120.

¹⁰³ The demodulation process was programmed using Hyperception's (2000) *Block Diagram*.

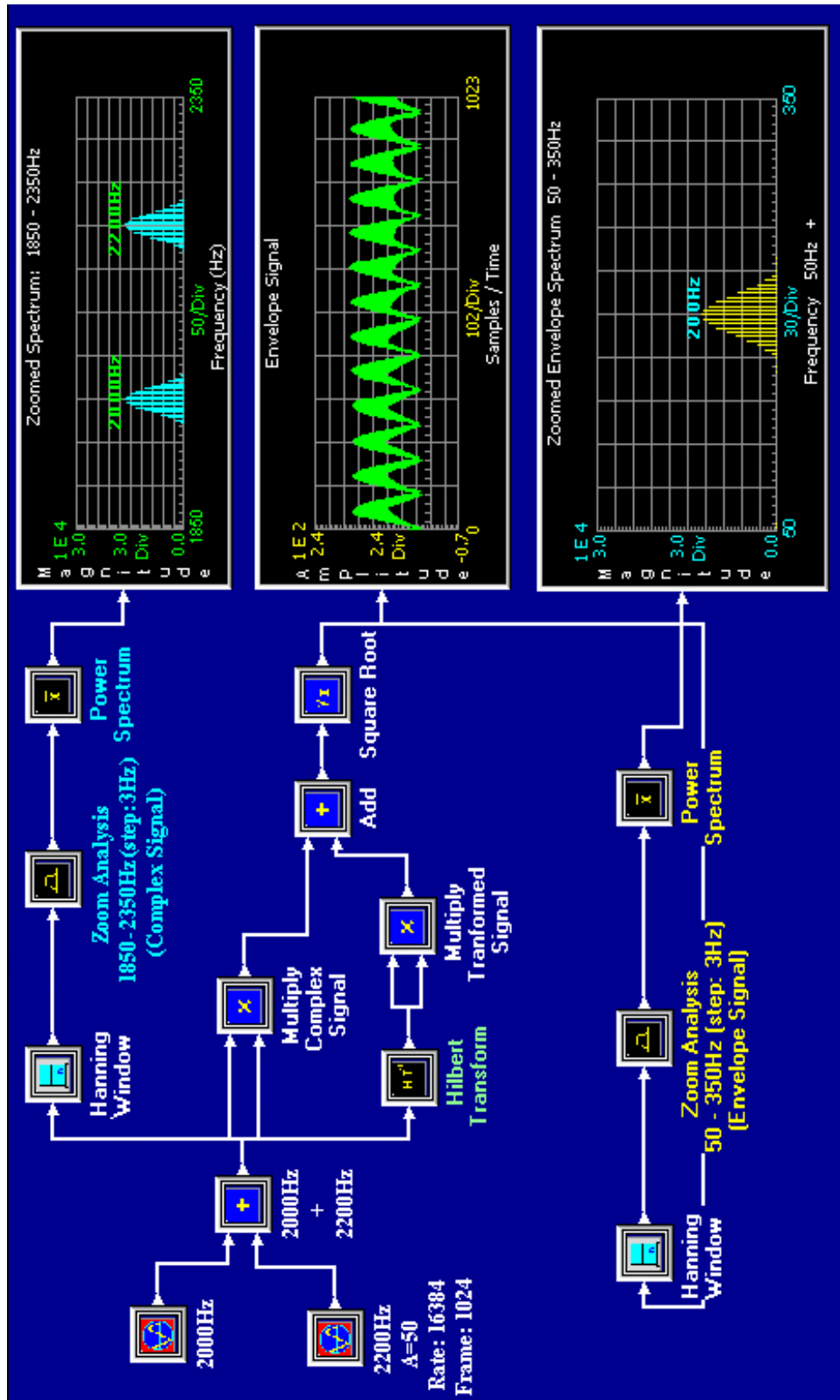


Figure 6.16: Block diagram of the Hilbert Transform Demodulation process extracting the envelope of a two-component complex signal and calculating its power spectrum (Experiment-4). Designed using Hyperception's (2000) *Block Diagram* software.

As illustrated in the block diagram (Figure 6.16), two cosine signals were created ($f_1 = 2000\text{Hz}$, $f_2 = 2200\text{Hz}$, $A_1 = A_2 = A = 50$, $\phi_1 = \phi_2 = 0$, *Sampling rate*: 16384 samples/sec, 16 bit, mono) and were added together resulting in a two-component complex signal of the type $y(t) = A(t) \cos(2\pi f t)$:

$$y(t) = A \cos 2\pi f_1 t + A \cos 2\pi f_2 t = 2A \cos 2\pi f_{\text{mod}} t (\cos 2\pi f_{\text{av}} t) \quad \text{Eq. (6.17)}$$

$$(f_{\text{av}} = \frac{f_1 + f_2}{2} = 2100\text{Hz}, f_{\text{mod}} = \frac{|f_2 - f_1|}{2} = 100\text{Hz}, \text{ and } f_{\text{AF}} = |f_2 - f_1| = 200\text{Hz})$$

The frequencies of the two components were chosen to satisfy the general condition:

$$\frac{f_1 + f_2}{2} = 2100\text{Hz} \gg \frac{|f_2 - f_1|}{2} = 100\text{Hz}$$

They were also chosen so that the period of the resulting amplitude fluctuations would be short enough ($T = 1/|f_2 - f_1| = 0.005 \text{ sec}$) to justify the assumption that the average power of the superposition over one period of the fluctuations is equal to the sum of the powers of the individual sines. Satisfying this assumption permitted the application of the energy conservation law.

Results

Frequency Zoom analysis¹⁰⁴ of the experiment signal around the frequencies of interest indicated the power spectrum in Figure 6.17. The frequencies with maximum energy were the expected frequencies of the two (interfering) components. The magnitude of each frequency component is proportional to amplitude.

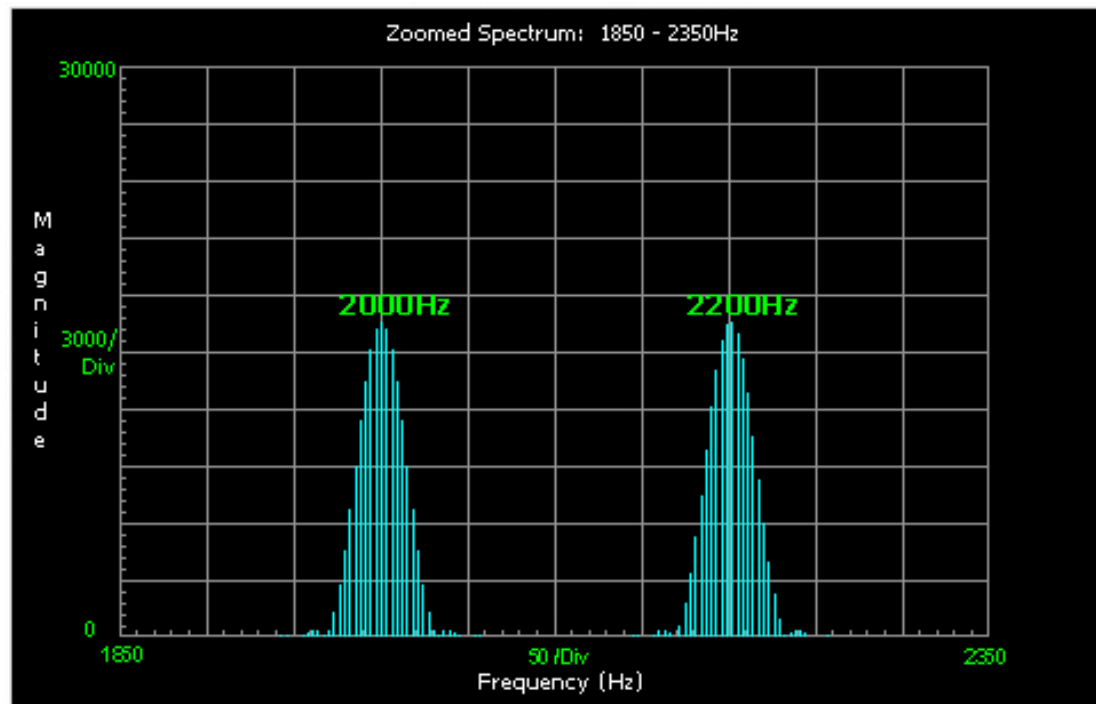


Figure 6.17: Frequency Zoom analysis of the signal in Experiment 4 (Hanning window). *Frequency range:* 1850Hz – 2350Hz; *Frequency Step:* 3Hz. Magnitude values are proportional to amplitude.

¹⁰⁴ *Frequency Zoom* analysis in Hyperception's (2000) *Block Diagram* uses the Goertzel algorithm to estimate the frequency amplitudes for a frequency range of interest. The Goertzel algorithm calculates the Discrete Fourier Transform (DFT) of a data set $y(n)$ using an Infinite Impulse Response filtering operation. The mathematical development of the algorithm is given in Proakis & Manolakis (1992: 737). Frequency Zoom analysis is inefficient relative to FFT when all frequency lines are required. However, it is very efficient for one or few frequency lines.

Performing the operation in Eq. (6.16), based on the Hilbert filter design shown in Figure 6.18, resulted in the envelope signal displayed in Figure 6.19.

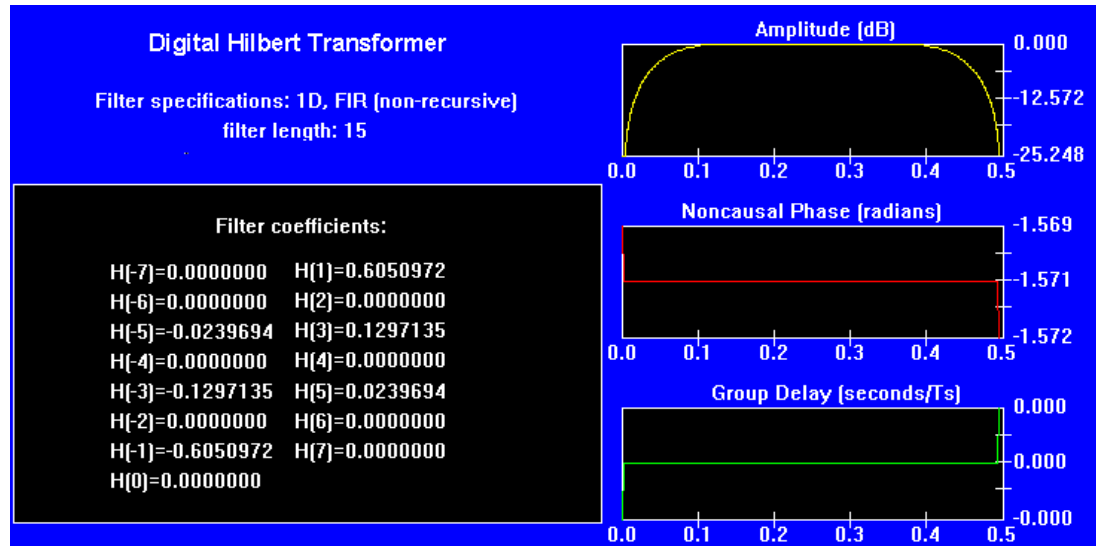


Figure 6.18: Hilbert Transform design for Experiment 4
(java applet design by Taft: <http://www.nauticom.net/www/jdtaft/hilbert.htm>).

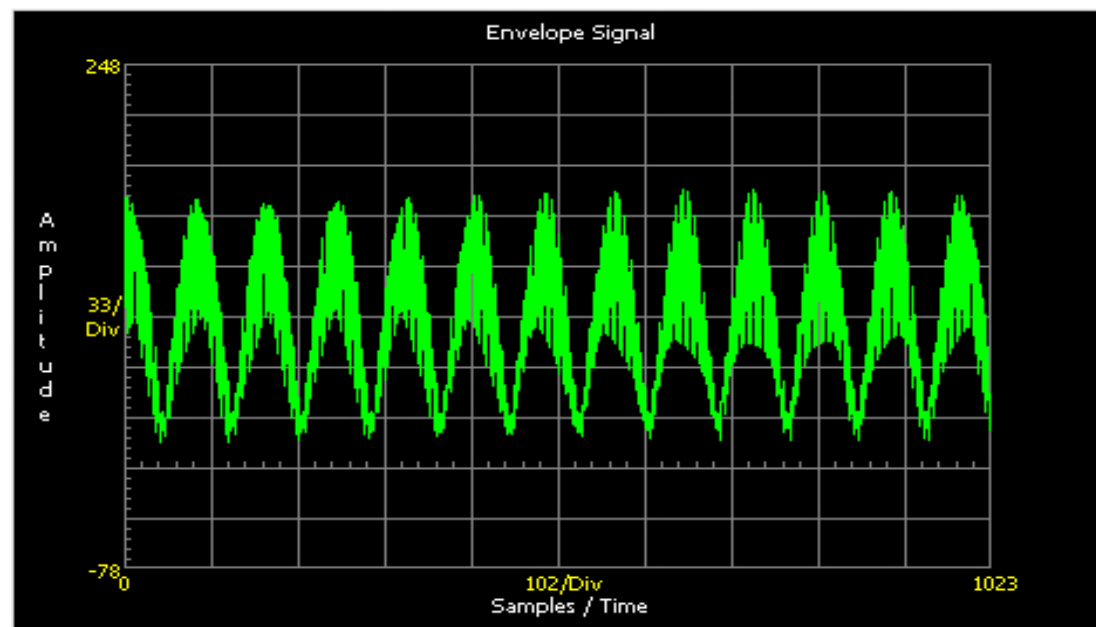


Figure 6.19: Envelope signal of the complex signal in Experiment 4, extracted using Eq. (6.16).

Frequency Zoom analysis of the envelope signal revealed a peak at 200Hz (Figure 6.20), a frequency value that matches the rate of amplitude fluctuation.

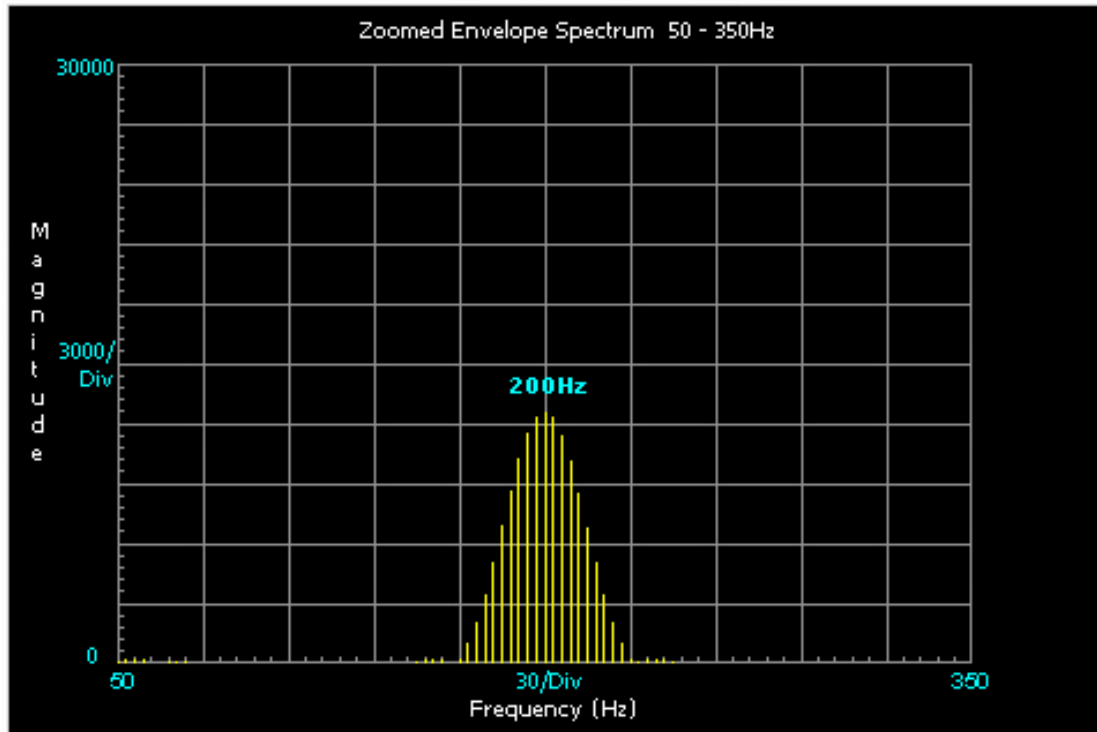


Figure 6.20: Frequency Zoom analysis of the envelope signal in Experiment 4 (Hanning window). *Frequency range:* 50Hz - 350Hz; *Frequency Step:* 3Hz. Magnitude values proportional to amplitude.

Application of a windowing filter during frequency analysis is often necessary to remove any ringing from the spectrum of a Hilbert transformed signal. As indicated in Figure 6.16, the Hanning window¹⁰⁵ was used for the analysis of both, the

¹⁰⁵ The *Hanning window*, also known as *Von Hann* or *cosine squared* window, has the following characteristics: *Frequency Roll* (frequency factor at which the amplitude is reduced by 3dB): 0.74; *Frequency Cut* (frequency factor at which the amplitude is equal to the maximum side-lobe attenuation: 34.7dB): 1.89; *Frequency Null* (frequency factor at which the first null amplitude is found): 2.05; *Normalization factor for unity gain* (0dB): window size/2 (Proakis & Manolakis, 1992).

transformed and the original signals, to ensure that the amplitude relationship among all identified components did not change due to windowing.

The sound synthesis / analysis frame size was 1024 samples. For the sampling rate used (16384 samples/sec), this corresponds to frequency resolution of 16Hz and temporal resolution of 62.5ms (period of 200Hz = 50ms.) Both Zoom analyses were in 3Hz steps. In order to compare the amplitude values of the interfering sines to the amplitude value of the fluctuations, five amplitude values were averaged from each analysis, centered on the frequencies of interest, and spanning a range (5x3=15Hz) that matched the analysis bandwidth (16Hz). Table 6.13 displays the results.

Table 6.13 Average Amplitude Values of the Interfering Sines and of the Amplitude Fluctuation Component in Experiment 4 Values from 5 bins of the frequency zoom analysis were averaged to span a range (15Hz) that matched the frequency resolution of the signal analysis (16Hz).			
Interfering Sines		Amplitude Fluctuation Component	
Frequency Bands	Amplitude: A	Frequency Bands	Amplitude: A_{AF}
<u>1994Hz(2194Hz)</u>	14402.88672	<u>194Hz</u>	11547.18652
<u>1997Hz(2197Hz)</u>	15862.54792	<u>197Hz</u>	12363.27246
<u>2000Hz(2200Hz)</u>	16538.84375	<u>200Hz</u>	12641.36914
<u>2003Hz(2203Hz)</u>	15936.44531	<u>203Hz</u>	12351.53418
<u>2006Hz(2206Hz)</u>	15146.33203	<u>206Hz</u>	11524.74023
Average: A	15577.41115	Average: A_{AF}	12085.62051
$A_{AF} = 0.776A$			

Based on the amplitude values in Table 6.13, envelope extraction and frequency analysis of the envelope signal indicated a frequency component at the rate of amplitude fluctuation with amplitude $A_{AF} = 0.776A$.

According to Eq. (6.6), the predicted amplitude for the amplitude fluctuation component is $A_{AF} = \sqrt{0.5}A = 0.707A < 0.776A$, indicating a difference between predicted and observed values. Some of the difference can be accounted for by the experimental method used to estimate the amplitude values of interest. Frequency analysis techniques introduce spectral spreading, or *leakage* that results from violating the periodicity and infinite-length assumptions for the data records analyzed. Moreover, the leakage is not symmetrical around each frequency component because of the leakage contributions of positive and negative frequencies into each other (Burgess, 1998). Windowing of the data reduces but does not eliminate leakage. Obtaining amplitude values by averaging over several analysis bands centered on a frequency of interest is a simplistic and imprecise leakage error-correction method. For accurate estimation of the amplitudes in a complex signal, Discrete Fourier Transform (DFT) analysis must be combined with more sophisticated leakage correction methods (in Burgess, 1998: 1073). In the present experiment, the analysis error has been augmented by expressing the amplitude value of interest (amplitude of the fluctuation component) as a function of another amplitude value (amplitude of one of the interfering sines) within the same analysis. The reason for expressing the

amplitude of the fluctuation component in relative rather than absolute terms is that the absolute amplitude level obtained from spectral analysis of an envelope signal has no meaning (Feldman, 1994). Nonetheless, the amplitude peaks in the spectrum of an envelope are often used as an indication of the envelope's power when compared to the mean noise floor. An example is the use of envelope extraction techniques to diagnose faults in mechanical gears, where the relative amplitudes in the envelope spectrum of a faulty gear's signal indicate a fault's severity (McLean et al., 1990). For this reason, although absolute amplitude levels of the envelope spectrum are meaningless, they may provide a measure of envelope power when seen relative to the levels of another component (within the same analysis) that has known power.

In conclusion, Hilbert demodulation of the complex signal in Experiment 4 and computation of the resulting envelope's power spectrum revealed a single component with frequency equal to the complex signal's amplitude fluctuation rate, providing clear evidence of the energy content of amplitude fluctuations. The amplitude value for the fluctuation component indicated by the analysis, $A_{AF} = 0.776A$, was different from the value predicted on the basis of the energy conservation law, $A_{AF} = 0.707A$ (Section 6.3.1). The discrepancy between predicted and observed amplitudes may be due to possible violations to the assumptions underlying the application of the energy conservation law, as well as to inadequate protection from analysis errors due to leakage. The experiment results regarding the amplitude of the fluctuation component are therefore inconclusive, indicating the need for further study.

6.3.4 Summary

It is argued that modulations in general and amplitude fluctuations in particular satisfy all criteria that define a wave. Section 6.3 investigated the energy content of amplitude fluctuations in an effort to provide theoretical and experimental support to the large amount of evidence demonstrating energy transfer at the rate of amplitude fluctuation:

- Amplitude fluctuations can transfer energy to systems (i.e. machinery, resonators, etc.) whose natural frequency matches the fluctuation rate.
- Amplitude fluctuations of acoustic waves correspond to changes in the vibration velocity of a mass at a rate equal to the rate of fluctuation, indicating consumption / radiation of energy at the same rate.
- Amplitude fluctuations entering a dispersive medium separate from the wave that carries them and travel independently at an observable rate and strength.
- Successive portions of a diffracted wave-front are determined by the envelope of secondary wavelets, indicating that amplitude fluctuations and the wave envelopes that describe them behave as waves.

A theoretical value for the energy content of amplitude fluctuations was calculated for the simplified case of a single point source oscillating at two angular

frequencies ω_1, ω_2 , with the same amplitudes $A_1 = A_2 = A$, and zero initial phases.

The calculation was based on applying the energy conservation law to Eq. (6.2).

$$y(t) = A \cos(\omega_1 t) + A \cos(\omega_2 t) = 2A \cos \omega_{\text{mod}} t (\cos \omega_{\text{av}} t) \quad \text{Eq. (6.2)}$$

$$\text{where} \quad \omega_{\text{av}} = \frac{\omega_2 + \omega_1}{2} \gg \omega_{\text{mod}} = \frac{|\omega_2 - \omega_1|}{2}.$$

Eq. (6.2) describes the two-component complex wave emanating from the oscillating point-source as a sine wave with fluctuating amplitude. Linearity of superposition does not apply to energy quantities of wave motion except under special cases of interference. In the simplified case of Eq. (6.2), the conservation law holds only under the following condition: The period of the amplitude fluctuations must be short enough to justify the assumption that the average power of the superposition over one period of the fluctuations is equal to the sum of the powers of the individual sines. Assuming that this condition and the condition in Eq. (6.2) are satisfied, the application of the energy conservation law indicates a power difference between the two sides of Eq. (6.2), interpreted as the power radiated by the fluctuations P_{AF} .

$$P_{AF} = \frac{\omega_{AF}^2 \rho A_{AF}^2}{8\pi c} \quad \text{Eq. (6.6)}$$

Eq (6.6) describes the power radiated by a point source pulsating with angular frequency $\omega_{AF} = |\omega_1 - \omega_2|$ (rate of amplitude fluctuation) and amplitude

$A_{AF} = \sqrt{0.5}A$, where ω_1 , ω_2 , and A are the angular frequencies and common amplitude of the interfering sines respectively.

An alternative sound signal representation has been proposed that confirms graphically the energy content of amplitude fluctuation predicted in Eq. (6.6). At the same time it solves the negative-amplitude and consistent-energy-representation problems associated with traditional two-dimensional signals. The new signal representation is based on the complex equation of motion and includes both the sine (imaginary) and cosine (real) terms describing a vibration. It results in spiral sine signals and twisted-spiral complex signals (Figures 4.1, 6.13, & 6.14 and the Appendix), similar to complex analytic signals. Three-dimensional signals illustrate that amplitude fluctuations and the signal envelopes that describe them are not just boundary curves but waves that trace changes in the total instantaneous energy of a signal over time, representing the oscillation of potential and kinetic energies within a wave. Three-dimensional spiral sine signals offer a consistent measure of the energy content of sine waves across frequencies, while twisted spiral complex signals account for the negative amplitude values predicted by the mathematical expression of amplitude fluctuation, map the parameters of amplitude fluctuation onto the twisting

parameters (Figure 6.14), and paint a more realistic picture of sound wave propagation in air.

In order to empirically verify the energy content of amplitude fluctuation, the frequency and amplitude values predicted by the theoretical analysis ($\omega_{AF} = |\omega_1 - \omega_2|$ and $A_{AF} = \sqrt{0.5}A$) were compared to values obtained through Hilbert demodulation of the complex signal in Eq. (6.2) and Frequency Zoom analysis of the resulting envelope signal.

The analysis revealed a single component with frequency equal to the complex signal's amplitude fluctuation rate, providing clear evidence of the energy content of amplitude fluctuation. However, the amplitude value of the fluctuation component indicated by the analysis, $A_{AF} = 0.776A$, was different from the predicted value, $A_{AF} = \sqrt{0.5}A = 0.707A$. The discrepancy between predicted and observed amplitudes may be attributed to possible violations of the assumptions underlying the application of the energy conservation law, as well as to inadequate protection from analysis errors due to leakage. The results confirm the energy content of amplitude fluctuation but are quantitatively inconclusive, indicating the need for further study.

6.4 Roughness Estimation Model – Roughness Degrees and Harmonic Interval Ratings in the Western Musical Tradition

In Chapter 2, examples drawn from three musical traditions demonstrated that manipulation of amplitude-fluctuation energy levels¹⁰⁶ and the associated changes in the sensation of roughness are significant to the creation of sonic variations in music cultures around the world. The present Section argues that, within the Western musical tradition where sensory roughness is on the whole avoided as dissonant, the consonance hierarchy of harmonic intervals in the equally-tempered chromatic scale corresponds to variations in roughness degrees.

6.4.1 Roughness Estimation Model for Complex Spectra

In order to test a hypothesis linking the sensation of roughness to Western musical tradition's concepts of consonance and dissonance, an appropriate roughness estimation model must be developed. As indicated in Section 5.3, existing roughness estimation models do not account for the roughness contribution of amplitude

¹⁰⁶ Through manipulation of fluctuation rate and degree.

fluctuation, they often fail to capture the effect of register on roughness, and demonstrate a relatively low degree of agreement between predicted and observed roughness levels. A model introduced by Sethares (1998) will serve as the basis for a new roughness-estimation model, which will include an appropriate variable to account for the relationship between roughness and degree of amplitude fluctuation.

Sethares' model offers the best theoretical fit to the observed relationships among the roughness of a sine-pair, the frequency separation of the two interfering sines, and register. The *roughness curves* in Figure 6.21 are based on a number of experimental studies examining the roughness of pairs of sines and illustrate the above-mentioned relationships:

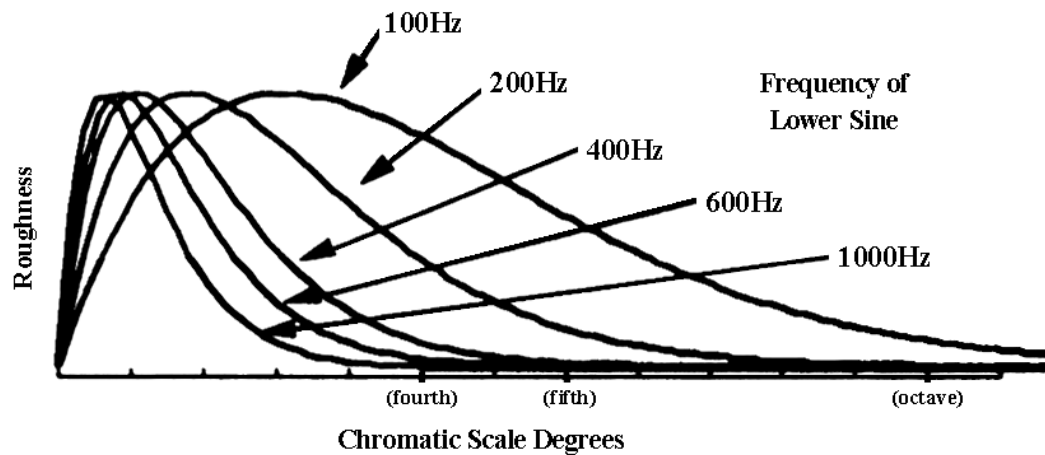


Figure 6.21: Roughness curves plotting the observed roughness (arbitrary measure - y axis) of a pair of sines (equal amplitudes) as a function of a) frequency separation (x axis) and b) frequency of the lower sine (after Sethares, 1998: 45). The roughness curves have been derived from multiple perceptual experiments examining the roughness of pairs of sines (Plomp & Levelt, 1965; Kameoka & Kuriyagawa, 1969a&b).

In Sethares' model, the roughness curves are interpreted as positively skewed Gaussian distributions:

$$y(x) = e^{-b_1 x} - e^{-b_2 x} \quad \text{Eq. (6.18)}$$

where x represents an arbitrary measure of the frequency separation $f_2 - f_1$, and b_1 & b_2 are the rates at which the function rises and falls.

Gradient minimization of the squared error between the experimental data (averaged over all frequencies) and the curve described by Eq. (6.18) gives: $b_1 = 3.5$ and $b_2 = 5.75$. For these values, the curve maximum occurs when $x = x^* = 0.24$, a quantity interpreted as representing the point of maximum roughness. To account for the observed nonlinearity in the relationship between the fluctuation rate corresponding to maximum roughness and the frequency of the lower sine (Figure 6.21), Sethares introduced the following modification, which includes the actual frequency spacing $f_2 - f_1$ and the frequency of the lower component f_1 to the calculation of roughness R :

$$R = e^{-b_1 s(f_2 - f_1)} - e^{-b_2 s(f_2 - f_1)} \quad \text{Eq. (6.19)}$$

$$\text{where } b_1 = 3.5, \ b_2 = 5.75, \ x^* = 0.24, \text{ and } s = \frac{x^*}{s_1 f_1 + s_2}.$$

The parameters s_1 and s_2 allow the function to stretch / contract with changes in the frequency of the lower component, so that the point of maximum roughness always agrees with the experimental data. A least-square fit gave $s_1 = 0.0207$ and $s_2 = 18.96$. To complete his model, Sethares multiplies Eq. (6.19) by the product of the amplitudes of the interfering sines, overestimating the contribution of sound pressure level (*SPL* - absolute amplitudes of the interfering sines) to roughness while underestimating the contribution of the degree of amplitude fluctuation (*AF-degree* - relative amplitudes of the interfering sines) (see Section 3.4).¹⁰⁷

For the new roughness estimation model, a term will be added to Eq. (6.19) that will account more reliably for the dependence of roughness on *SPL* and *AF-degree*. Terhardt (1974a) examined this dependence experimentally and provided a set of related functions. These functions had to be adjusted (Chapter 3) to reflect the quantitative difference between amplitude modulation depth (used to manipulate Terhardt's stimuli) and degree of amplitude fluctuation (needed in the roughness model).

The adjusted functions (see Section 3.4) are:

a) Roughness R of a three-component AM tone as a function of *AF-degree*:

¹⁰⁷ This is the case with most existing roughness estimation models (i.e. Helmholtz, 1885; Plomp & Levelt, 1965; Kameoka & Kuriyagawa, 1969a&b; Hutchinson & Knopoff, 1978).

$$R = cAF_{degree}^{3.11} \quad (c: \text{constant}) \quad \text{Eq. (6.20)}$$

b) Roughness R of a three-component AM tone as a function of SPL :

$$R = cSPL^{\frac{1}{10}} \quad (c: \text{constant}) \quad \text{Eq. (6.21)}$$

c) Roughness R_{pair} of a sine-pair $(f_1, f_2; A_1, A_2)$ as a function of the roughness

R_{AM} of a three-component AM tone with the same amplitude fluctuation rate

$$(f_{mod} = |f_1 - f_2|) \text{ and degree } (AF_{degree} = \frac{2A_2}{A_1 + A_2} \Rightarrow AM_{depth} = m = \frac{A_2}{A_1}, A_1 \geq A_2):$$

$$R_{pair} = 0.5R_{AM} \quad \text{Eq. (6.22)}$$

Combining Eqs. (6.19), (6.20), (6.21), & (6.22) results in a new model for the estimation of the roughness R of pairs of sines with frequencies f_1 & f_2 , amplitudes A_1 & A_2 ($A_1 \geq A_2$) and equal (i.e. 0) initial phases:

$$R(f_1, f_2, A_1, A_2) = (A_1 * A_2)^{0.1} * 0.5 \left(\frac{2A_2}{A_1 + A_2} \right)^{3.11} [e^{-b_1 s(f_2 - f_1)} - e^{-b_2 s(f_2 - f_1)}] \quad \text{Eq. (6.23)}^{108}$$

¹⁰⁸ As a reminder, the degree of amplitude fluctuation (AF_{degree}) of a two component complex signal $(f_1, A_1; f_2, A_2; |f_1 - f_2|/2 \ll |f_1 + f_2|/2; A_1 \geq A_2)$ is: $AF_{degree} = 2A_2/(A_1 + A_2)$.

where $b_1 = 3.5$, $b_2 = 5.75$, $s = \frac{x^*}{s_1 f_1 + s_2}$, $x^* = 0.24$, $s_1 = 0.0207$, & $s_2 = 18.96$.

As discussed in Section 5.3.2, the roughness of complex spectra with more than two sine components will be estimated by adding together the roughness of the individual sine-pairs (Terhardt, 1974a; Lin & Hartmann, 1995).

6.4.2 Limitations of the Proposed Roughness Estimation Model

Three limitations of the proposed model are identified regarding phase, envelope asymmetry, and continuous spectral distributions, outlining the conditions for its application.

6.4.2.1 Effect of Phase on Degree of Amplitude Fluctuation

According to a study by Pressnitzer & McAdams (1999), when more than two sine components fall within the same critical band, their relative phase influences the total roughness. The authors indicate that the dependence of roughness on the phase is

not due to the phase values themselves but to the way these values affect the shape of the resulting signal's amplitude fluctuations.¹⁰⁹

Depending on the phase relationship of the interfering sines, AM-depth =100% may result in AF-degree <100% (Figure 6.22), accompanied by a decrease in the perceived roughness. This finding is consistent with the present study's argument on the difference between amplitude modulation depth (referring to the spectral distribution of a signal) and degree of amplitude fluctuation (referring to the envelope shape of a signal), linking the latter to roughness. Figure 6.22 illustrates the effect of phase on the degree of amplitude fluctuation for a fixed modulation depth (100%).

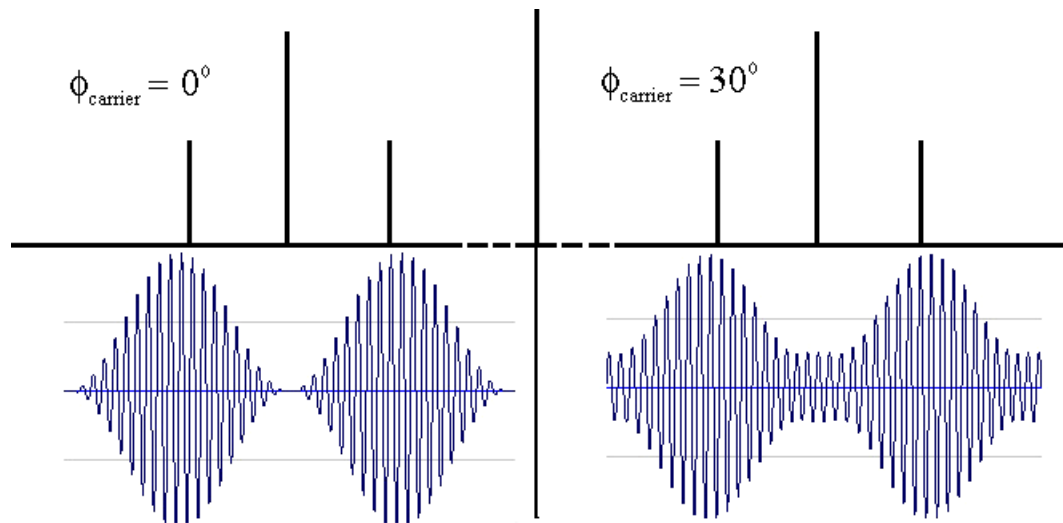


Figure 6.22: Effect of phase on the degree of amplitude fluctuation. Two 3-component complex signals with the same amplitude modulation depth (100%) but different degrees of amplitude fluctuation, resulting from the phase relationships between carrier and sidebands in the two spectra (the phase of the sidebands is assumed to be fixed: $\phi_{\text{left}} = \phi_{\text{right}} = 0^\circ$).
After Pressnitzer & McAdams, 1999.

¹⁰⁹ As the authors noted, Mathes & Miller made a similar observation in 1947.

The proposed roughness estimation model assumes zero initial phases for all interfering sines and does not capture the effect illustrated in Figure 6.22.¹¹⁰

Therefore, one condition for the application of the model is that the complex spectra examined have components with the same initial phase *or* less than three components per critical band.

It should be noted that the amplitude envelopes are the same for $\phi_{carrier} = -30^\circ$ and $\phi_{carrier} = 30^\circ$. Subjects, however, reported more roughness for the signals with positive carrier phase. Pressnitzer & McAdams (1999) explained the observation by claiming an effective asymmetry of the signal's envelope at a neural level, caused by the hypothesized characteristics of the auditory filter. This explanation agrees with their results on roughness ratings of asymmetrical amplitude envelopes (see Section 6.4.2.2). However, this explanation requires that the effective amplitude of the low-frequency sideband is lower than the amplitude of the high-frequency sideband. This requirement can be satisfied only if the center frequency of the auditory filter is different from the carrier frequency of the modulated signal, a condition equivalent to the auditory system resembling a set of fixed parallel filters. Nonetheless, the authors

¹¹⁰ In order for the proposed model to account for this effect, the following steps need to be taken: a) A function must be introduced to describe the relationship between phase and AF-degree. b) A perceptual experiment must be conducted examining the roughness relationship between AM tones with a specified AF-degree and AM tones with the same AF-degree arrived at through phase manipulation. c) A coefficient must be added to Eq. (6.23) implementing the results from steps (a) and (b).

adopted their explanation instead of an explanation by Buunen et al. (1974), which was based on the difference in the aural combination tones resulting from carriers with positive or negative phase. The explanation by Buunen et al. was dismissed because the between-subjects variance was so high that individual data had to be presented to report the effect. However, the case was not that different with the data of Pressnitzer & McAdams. Because of the experimental design, different outcomes were expected for the same roughness comparison across listeners or with repeated comparisons for the same listener. Pressnitzer & McAdams' data was more consistent because variance was reduced through the use of the Bradley-Terry-Luce (BTL) data scaling method.¹¹¹ Buunen et al. did not use this method. Regarding the observed roughness difference among signals with asymmetrical envelopes (see Section 6.4.2.2), an explanation other than the currently available ones needs to be introduced. Such an explanation will have to be consistent with the observed loudness differences among signals with asymmetrical envelopes, which are opposite to the observed roughness differences (Stecker & Hafter, 2000).

¹¹¹ The Bradley-Terry-Luce (BTL) data scaling method constructs a psychophysical scale out of binary-paired comparison judgments, reducing the overall variance.

6.4.2.2 Effect of Envelope Asymmetry on Roughness

Signals with the same degree of amplitude fluctuation and the same envelope rms values (i.e. same area outlined by the envelope and the time axis on the signal graph) may result in different roughness levels depending on the actual envelope shape (Pressnitzer & McAdams 1999). More specifically, for temporally asymmetrical amplitude envelopes (i.e. sawtooth, reversed sawtooth envelopes, etc.), signals whose amplitude rises faster than it drops (Figure 6.23(a)) sound rougher than signals whose amplitude rises slower than it drops (Figure 6.23(b)). Roughness differences increase with asymmetry and small departures from symmetry result in negligible roughness differences.¹¹²

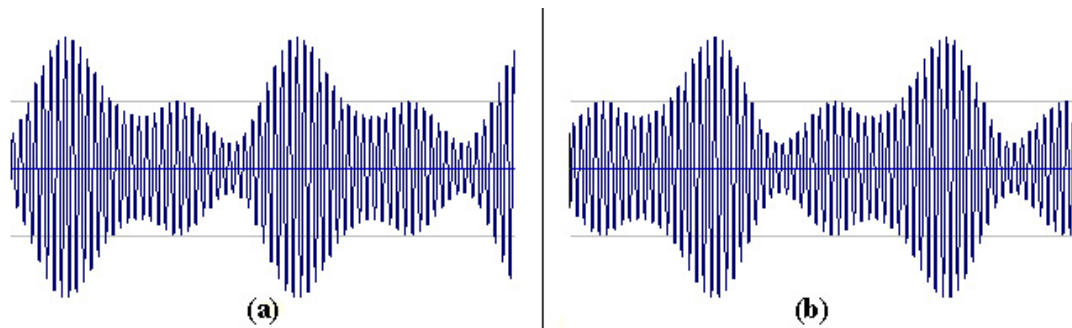


Figure 6.23: Effect of envelope asymmetry on roughness. The signal in (a) is judged as rougher than the signal in (b) although the rate, degree, and rms of the amplitude fluctuations are the same (after Pressnitzer & McAdams, 1999).

¹¹² The effect has not been examined systematically yet.

Stecker & Hafter (2000) found the opposite to be the case with loudness. As they noted, their results are compatible neither with automatic-gain-control, power-spectrum models of loudness, nor with auditory-image models. More studies are necessary before a satisfactory explanation that accounts for both asymmetries can be found.

The proposed roughness estimation model does not account for the effect of envelope asymmetry on roughness.¹¹³ Therefore, a second condition for the application of the model is that the complex spectra examined correspond to signals that are relatively symmetrical along the time axis.

¹¹³ In order for the proposed model to account for this effect, the following steps need to be taken: a) A function must be introduced to quantify envelope asymmetry b) A perceptual experiment must be conducted to examine systematically the effect of envelope asymmetry on roughness. c) A coefficient must be added to Eq. (6.23) implementing the results of steps (a) and (b).

6.4.2.3 Roughness of Continuous spectra

The proposed model is not efficient for signals with continuous spectra (i.e. noise). For such signals, it must be combined with loudness estimation and masking models, and must have frequency bands rather than frequency components as inputs. Daniel & Weber (1997) and Leman (2000) have proposed two such models. The Daniel & Weber model underestimates the contribution of the degree of amplitude fluctuation to the estimated roughness. As a result, it captures well the perceived low roughness of white noise and band-pass noise but is not as successful with periodic signals. Leman (2000) compares a cochlear model of roughness to his *Synchronization Index Model*, showing a good agreement between approaches. There are no perceptual data for either approach. Additionally, the cochlear model used in the study is relatively crude in its representation of the dependence of roughness on register, indicating that this may be the case with the *Synchronization Index Model* as well.

Therefore, in spite of its inability to handle continuous spectral distributions, the proposed model will be adopted conditionally¹¹⁴ because it represents more accurately the effect of AF-degree and register on roughness.

¹¹⁴ The spectra examined must be discrete.

6.4.2.4 Summary

Section 6.4.2 identified three limitations of the proposed roughness estimation model. The model does not capture the effect of phase or of the precise amplitude envelope shape on roughness, and can only be applied to signals with discrete spectra. To ensure the reliability of the model, the complex spectra examined must a) have components with the same initial phase *or* have less than three components per critical band, b) correspond to signals that are relatively symmetrical along the time axis, and c) be discrete. Under these conditions, the currently proposed model (Eq. 6.23)) constitutes a good representation of the theoretical knowledge and empirical data on the roughness of complex spectra. The following Section applies the model to an empirical examination of the relationship between roughness degrees and harmonic interval consonance ratings in the Western musical tradition.

6.4.3 Experiment 5: Roughness Degrees and Harmonic Interval Dissonance

Ratings

Hypothesis

For musicians within the Western musical tradition, roughness ratings of harmonic intervals agree with the roughness degrees estimated using the proposed roughness estimation model (Eq. (6.23))¹¹⁵ and correlate with the dissonance degrees suggested by Western music theory. Additionally, dissonance ratings correlate with roughness degrees indicating that, in the Western musical tradition where sensory roughness is in general avoided as dissonant, the consonance hierarchy of harmonic intervals corresponds mainly to variations in roughness degrees.

Helmholtz (1885), Pratt (1921), and others (see Section 5.3) have conducted somehow similar experiments that compared roughness model estimations or roughness ratings to music theory claims. However, no study to date has examined simultaneously the relationship among roughness model estimations, roughness ratings, dissonance ratings, and music theory claims.

¹¹⁵ The spectra of the experiment stimuli satisfied the conditions outlined in Section 6.4.2, warranting the use of the proposed roughness estimation model.

Method

The thirteen harmonic intervals of the chromatic scale starting on middle C (C4; fundamental frequency: 256Hz; equal temperament) served as experiment stimuli. The intervals were constructed using synthesized complex tones with six components each and static, sawtooth spectra. The frequency components were slightly shifted away from a harmonic relationship. The sawtooth and slightly detuned spectra were chosen to introduce 'naturalness' to the stimuli. Based on results from previous studies (von Békésy, 1960; Terhardt, 1974a), the low frequency / low-level beating caused by the detuning was not expected to influence the roughness ratings.

Along with the predictions of the proposed model (Eq. (6.23)), the predictions of two additional roughness estimation models (Helmholtz, 1885: 332; Hutchinson & Knopoff, 1978: 17-23) were examined for comparison purposes. To facilitate comparison, predicted roughness values were scaled so that all roughness estimation models spanned the same roughness range. All models assume sawtooth spectra ($A_n = A_1/n$; A_n : amplitude of the n^{th} component) and the same starting note, C4. Both earlier models assume ten- rather than six-component complex tones. The Hutchinson & Knopoff model is based on Helmholtz's model, modified to reflect the results by Plomp & Levelt (1965) regarding the effect of register on roughness. Plomp & Levelt examined this important effect using six-component complex tones. Therefore, the present study opted for the use of six- rather than ten-component complex tones.

Table 6.14 displays the frequency (in Hz) and amplitude (in arbitrary units) values of the components of the complex tones used to construct the thirteen harmonic intervals.

Table 6.14 Frequency (in Hz) and Amplitude (in arbitrary units) Values of the Components of the Complex Tones Used to Construct the Stimuli in Experiment 5 The 13 harmonic intervals of the chromatic scale (starting on C4) were the stimuli for Experiment 5, testing the relationship among estimated roughness, roughness ratings, and dissonance ratings.						
<u>Component</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>
<u>Amplitude</u>	<u>1</u>	<u>0.5</u>	<u>0.333</u>	<u>0.25</u>	<u>0.2</u>	<u>0.166</u>
<u>Note</u>						
<u>C4</u>	262	526	790	1049	1318	1573
<u>C#4 / Db4</u>	277	554	837	1118	1398	1677
<u>D4</u>	294	590	886	1180	1473	1772
<u>D#4 / Eb4</u>	311	624	932	1244	1569	1873
<u>E4</u>	330	663	995	1323	1654	1994
<u>F4</u>	349	701	1053	1408	1751	2107
<u>F#4 / Gb4</u>	370	741	1118	1482	1852	2235
<u>G4</u>	392	783	1179	1570	1973	2373
<u>G#4 / Ab4</u>	415	834	1250	1670	2093	2499
<u>A4</u>	440	884	1329	1768	2200	2666
<u>A#4 / Bb4</u>	466	937	1400	1874	2345	2799
<u>B4</u>	494	990	1484	1985	2476	2973
<u>C5</u>	524	1052	1573	2110	2634	3154

Intervals were presented randomly to two groups of ten subjects through headphones (same signal in both ears). The first group of subjects was asked to rate the stimuli in terms of roughness, on a scale outlined by the labels: *Not rough - Rough*. The second group of subjects was asked to rate the stimuli in terms of dissonance, on a scale outlined by the labels: *Not dissonant - Dissonant*. Subjects were able to familiarize themselves with the stimuli in a practice run and demonstrated their understanding of *roughness* in training sessions that included amplitude-modulated sines as stimuli, at various modulation rates and depths.

Response scales ranged from 0 (Not Rough / Not Dissonant) to 42 (Rough / Dissonant). The scroll-bar starting position was random. The range (0-42) was based on the roughness of the stimuli calculated using a computer implementation of the proposed model, Eq. (6.23). The predictions of the two comparison models were scaled to fit this range.

Results

Table 6.15 displays the mean responses of two groups of ten subjects, rating the experiment stimuli in terms of roughness and dissonance respectively (Columns 3 & 4). The gray cells under the second column indicate the estimated roughness of each stimulus based on Helmholtz (1885: 332), Hutchinson & Knopoff (1978: 17-23), and the proposed model, Eq. (6.23).

Table 6.15 Mean Roughness and Dissonance Ratings for 13 Harmonic Intervals versus roughness estimated by 3 roughness estimation models (2 Groups of 10 Subjects; Experiment 5) The experiment examined the relationship among estimated roughness, observed roughness, and observed dissonance for two existing roughness estimation models (Helmholtz, 1885; Hutchinson & Knopoff, 1978) and for the proposed model (Eq. (6.23)) .						
<u>Interval</u>		<u>Estimated Roughness</u>			<u>Observed Roughness</u>	<u>Observed Dissonance</u>
		<u>Helmholtz</u>	<u>Hutchinson & Knopoff</u>	<u>Eq. (6.23)</u>		
Unison	C4 - C4	0	0.157	1.614	1.7	0.8
Minor 2 nd	C4 - C#4	40.383	40.383	40.383	36.6	38.2
Major 2 nd	C4 - D4	13.284	22.233	27.617	29.7	36.4
Minor 3 rd	C4 - Eb4	12.753	9.166	18.117	19.2	15.5
Major 3 rd	C4 - E4	9.564	4.554	16.002	14.3	10.1
Fourth	C4 - F4	1.594	3.727	11.446	12.5	6.6
Aug. 4 th	C4 - F#4	9.564	7.686	12.826	13	21.4
Fifth	C4 - G4	0.531	1.826	6.17877	4.7	4.4
Minor 6 th	C4 - Ab4	11.69	6.967	10.103	10.7	15
Sixth	C4 - A4	11.69	3.942	5.782	8.2	5.1
Minor 7 th	C4 - Bb4	12.753	8.249	6.214	10.4	13.5
Major 7 th	C4 - B4	25.506	19.109	6.996	9.2	12
Octave	C4 - C5	0	0.116	1.589	2.5	1.7

Figure 6.24 displays the estimated roughness for the thirteen harmonic intervals used in Experiment 5, based on three roughness estimation models.

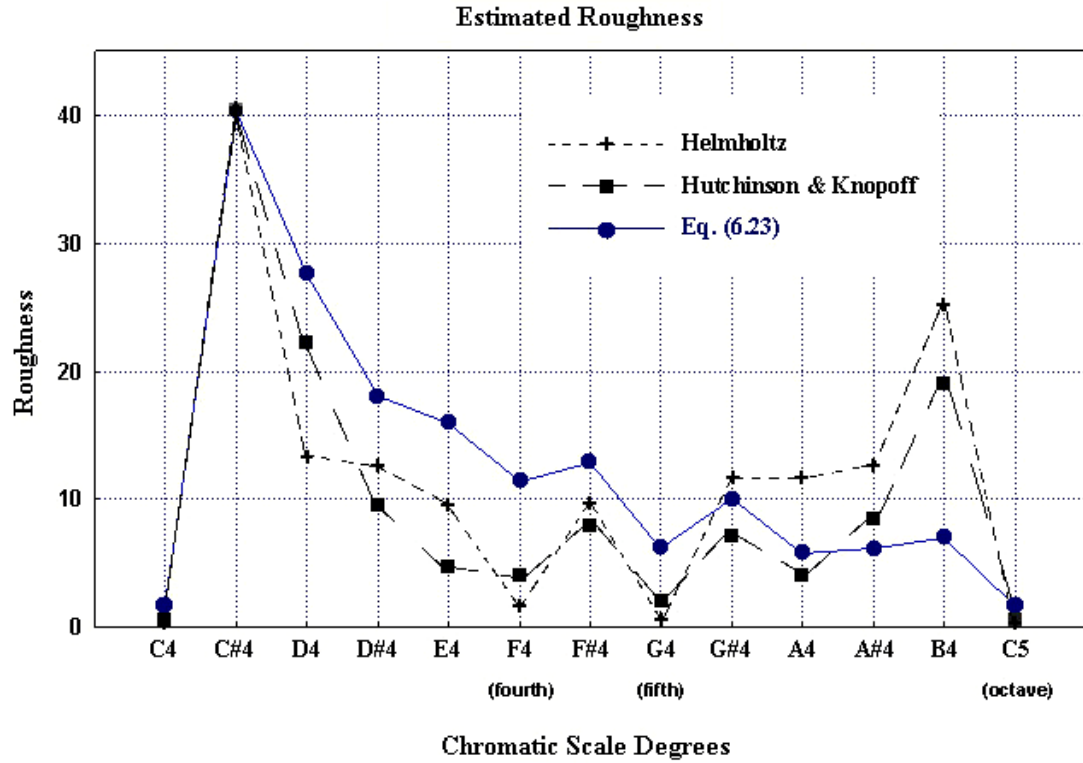


Figure 6.24: Estimated roughness for all intervals 13 in the chromatic scale starting at C4 (Experiment 5). Values were estimated using the proposed model (Eq. (6.23)) and two existing roughness estimation models: Helmholtz, 1885 and Hutchinson & Knopoff, 1978.

The roughness estimates of the proposed model correlate better with the estimates of the Hutchinson & Knopoff model ($r = 0.86$) than with those of the Helmholtz model ($r = 0.72$). The main reason for this difference is that the Hutchinson & Knopoff model accounts for the effect of register on roughness while the Helmholtz model does not.

Based on Figure 6.24, a number of differences in the roughness predictions of the three models can be identified:

- a) The proposed model (Eq. (6.23)) predicts a much lower roughness level for the Major 7th interval (C4-B4) than the other two models.
- b) The Helmholtz and the Hutchinson & Knopoff models predict lower roughness levels for the intervals between Major 2nd (C4-D4) and Fifth (C4-G4).
- c) According to the proposed model (Eq. (6.23)), the Augmented 4th interval (C4-F#4) is smoother than the Major 3rd (C4-E4).
- d) Eq. (6.23) results in a roughness curve with a relatively linear shape between the Minor 2nd (C4-C#4) and Octave (C4-C5) intervals, without the pronounced contrasts found in the roughness curves of the other models.
- e) Lastly, the proposed model predicts slightly higher roughness levels for the Unison (C4-C4) and Octave (C4-C5) intervals than the other models.

The slightly higher roughness levels predicted for the Unison and Octave intervals by the proposed model (Eq. (6.23)) are due to the slight detuning applied to the experimental stimuli. The Helmholtz and the Hutchinson & Knopoff models assume perfectly harmonic spectra that bring the roughness levels of Unisons and Octaves very close to zero.

All other differences can be explained in terms of the different assumptions each model makes regarding the contributions of degree of amplitude fluctuation (*AF-degree*) and of sound pressure level (*SPL*) to roughness. The proposed model assumes the power relationships outlined in Eq. (6.20) ($R = cAF_{degree}^{3.11}$) and Eq. (6.21) ($R = cSPL^{\frac{1}{10}}$). The Helmholtz model ignores the contribution of *AF-degree* and assumes a linear relationship between *SPL* and roughness. The Hutchinson & Knopoff model assumes a linear relationship to roughness for both, *AM-depth* (used to manipulate *AF-degree*) and *SPL*.

The unwarranted assumptions of the last two models result in overestimating the roughness contribution of some sine-pairs, while underestimating the roughness contribution of others.

For the Major 7th interval (C4-B4), for example, the roughness from the interaction between the 2nd, 4th, and 6th components of C4 and the 1st, 2nd, and 3rd components of B4 respectively¹¹⁶ has been overestimated. All three pairs have sines with largely unequal amplitudes that result in low *AF-degree* and therefore lower roughness than estimated by both the Helmholtz and the Hutchinson & Knopoff models.

For the Fourth interval (C4-F4), the roughness from the interaction between the 3rd, 4th, and 5th components of C4 and the 2nd, 3rd, and 4th components of F4

¹¹⁶ These three sine-pairs contribute the most to the roughness of the interval because of their amplitude fluctuation rates (i.e. frequency differences of the sines within each pair).

respectively has been underestimated. The sines in these three pairs have similar amplitudes resulting in high *AF-degrees* and higher roughness than estimated by the Helmholtz and the Hutchinson & Knopoff models. The higher roughness of the Fourth interval relative to the Major 7th, predicted by the proposed model, is consistent with experiments indicating that Fourths are less likely to be perceived as single tones (less prone to tonal fusion) than Major 7ths (DeWitt & Crowder, 1987, in Huron, 1991: 136).

The difference in the roughness ranking of the Augmented 4th (C4-F#4) among the three models can be explained in similar terms.

The somewhat linear shape of the proposed model's roughness curve is due to the progressive decrease in *AF-degree*¹¹⁷ for the sine-pairs that contribute the most to the roughness of each interval.

Figures 6.25 and 6.26 display the mean responses of the subject-groups rating the stimuli in terms of roughness and dissonance respectively. Figure 6.27 combines the results.

¹¹⁷ As the intervals get wider, the amplitude difference of the closely interacting sine components gets larger and the *AF-degree* gets smaller.

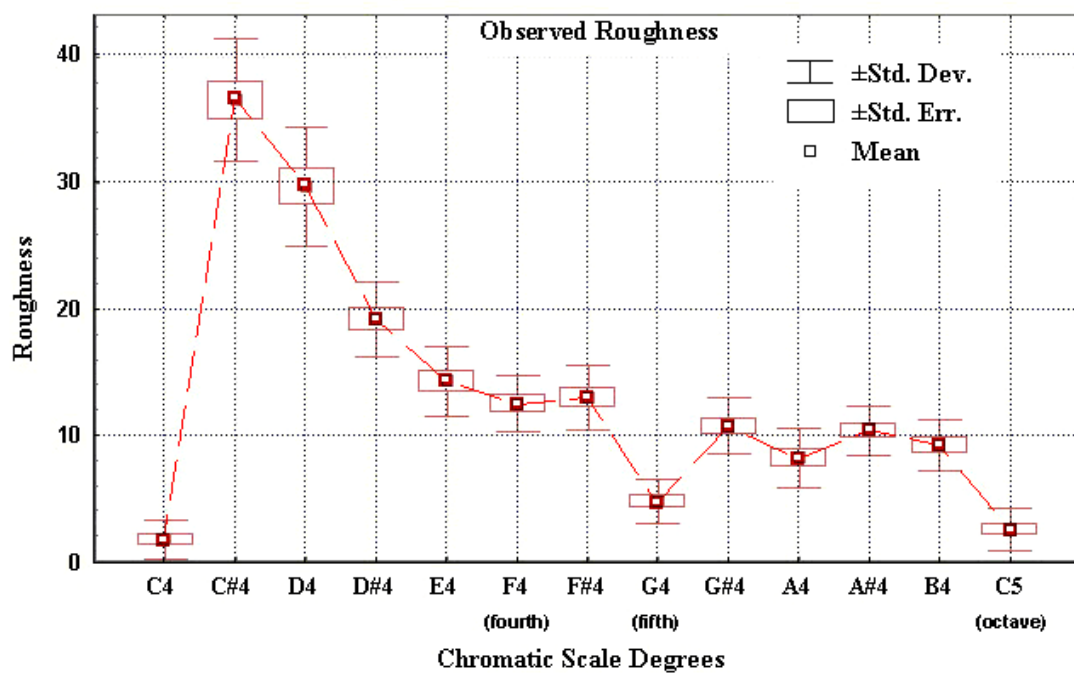


Figure 6.25: Mean, standard deviation, and standard error values (10 subjects) of the observed roughness for all 13 intervals in the chromatic scale starting at C4 (Experiment 5).

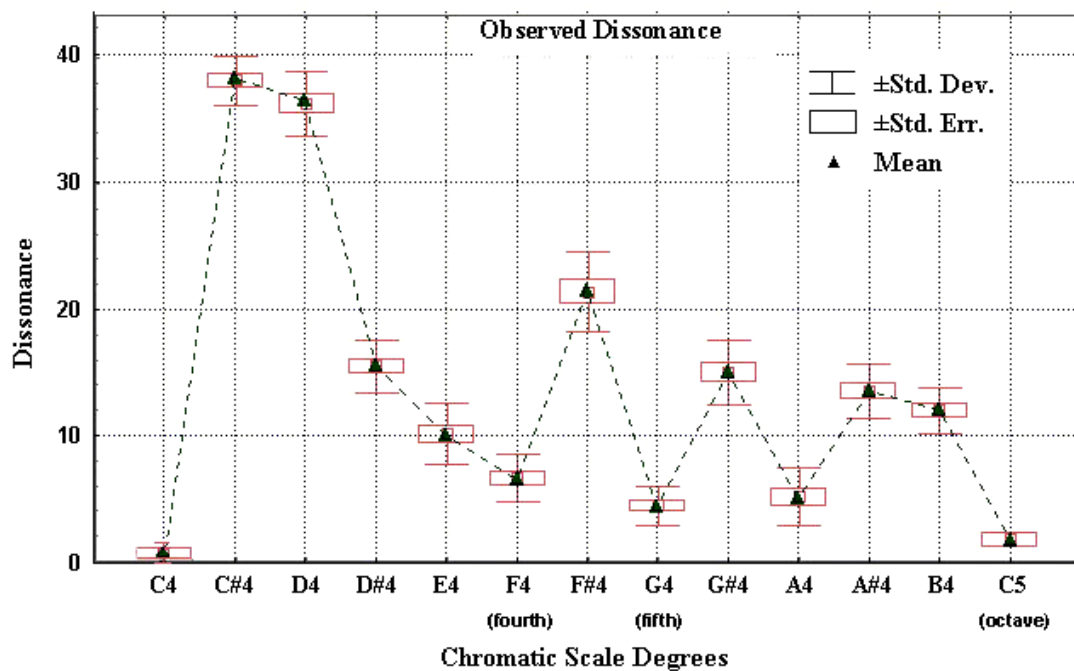


Figure 6.26: Mean, standard deviation, and standard error values (10 subjects) of the observed dissonance for all 13 intervals in the chromatic scale starting at C4 (Experiment 5).

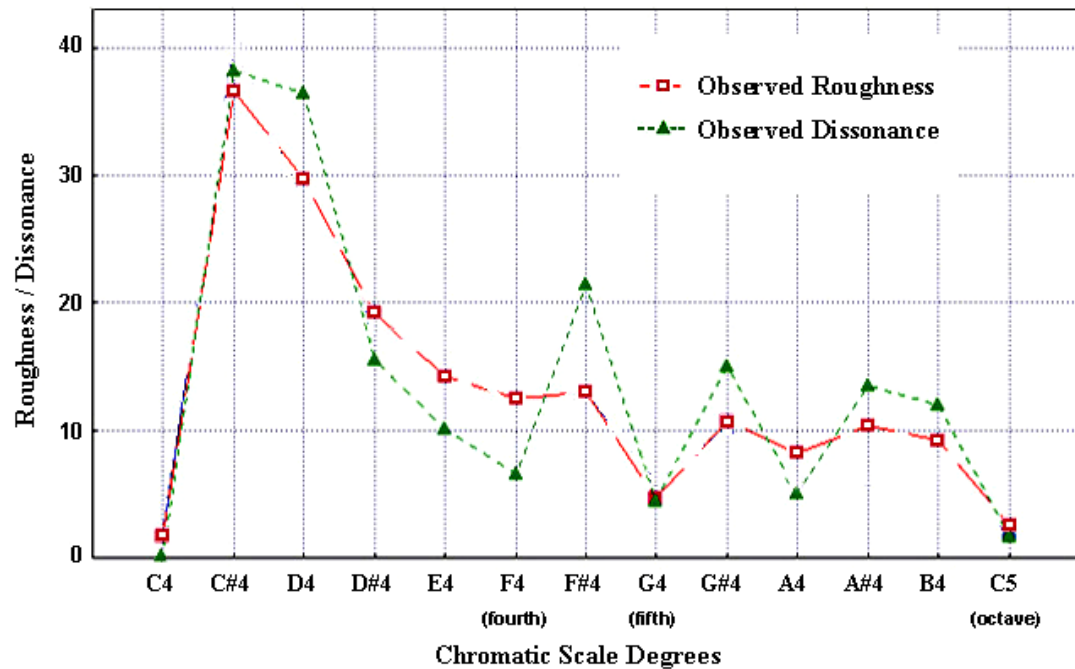


Figure 6.27: Mean observed roughness versus mean observed dissonance (10 subjects per group) for all 13 intervals in the chromatic scale starting at C4 (Experiment 5).

Dissonance responses correlated well with roughness responses ($r = 0.94$), indicating that changes in dissonance among the thirteen intervals corresponded to changes in roughness. This suggests that the presence of roughness provides an important clue for dissonance judgments of isolated harmonic intervals.

Analysis of variance¹¹⁸ suggested that the subjects rating dissonance used slightly different criteria than the subjects rating roughness ($F(1,18) = 4.46$; $p = 0.0489$). Post-hoc analysis (Scheffé test; $\alpha = 0.05$) indicated that this difference was manifested the strongest in the ratings of the Augmented 4th ($F(1,18) = 42.22$; $p < 0.0001$), the Fourth ($F(1,18) = 41.83$; $p < 0.0001$), and the Minor 6th ($F(1,18) = 15.98$; $p = 0.0008$) intervals. The Fourth interval was judged less dissonant than rough while the Augmented 4th and Minor 6th intervals were judged more dissonant than rough.

To further investigate the observed differences, two separate analyses were performed on the means obtained from the roughness and the dissonance subject groups, organizing the intervals into the roughness and dissonance categories displayed in Table 6.16. The analysis used the Tuckey HSD instead of the Scheffé post-hoc test ($\alpha = 0.05$). Due to the small number of subjects and the conservative nature of the Scheffé test, initial analysis by Scheffé revealed very few differences between means. For Experiment 5, the Tuckey HSD test provided adequate protection against the increased *alpha* error rate due to multiple post-hoc comparisons, while providing results that were easier to interpret.¹¹⁹

¹¹⁸ The design had one between-subjects fixed factor (*Rating*) with two levels (roughness / dissonance) and one within subjects random factor (*Interval*) with thirteen levels (thirteen harmonic intervals in the chromatic scale).

¹¹⁹ The Tuckey HSD test, although less conservative (or because of it), is in general more widely used than the Scheffé test.

Table 6.16 Mean Roughness and Dissonance Ratings Grouped According to Statistical Significance by Tuckey HSD Post-Hoc Comparisons ($\alpha = 0.05$); Experiment 5					
Roughness			Dissonance		
Category	Interval	Mean Roughness	Category	Group 4	Mean Dissonance
1	<u>Unison</u>	1.7	1	<u>Unison</u>	0.8
1	<u>Octave</u>	2.5	1 2	<u>Octave</u>	1.7
1 2	<u>Fifth</u>	4.7	2 3	<u>Fifth</u>	4.4
2 3	<u>Major 6th</u>	8.2	3	<u>Major 6th</u>	5.1
3 4	Major 7 th	9.2	3	Fourth	6.6
3 4 5	Minor 7 th	10.4	4	Major 3 rd	10.1
3 4 5	Minor 6 th	10.7	4 5	Major 7 th	12
4 5	Fourth	12.5	5 6	Minor 7 th	13.5
4 5	Aug. 4 th	13	5 6	Minor 6 th	15
5	Major 3 rd	14.3	6	Minor 3 rd	15.5
6	Minor 3 rd	19.2	7	Aug. 4 th	21.4
7	<u>Major 2nd</u>	29.7	8	<u>Major 2nd</u>	36.4
8	<u>Minor 2nd</u>	36.6	8	<u>Minor 2nd</u>	38.2

Table 6.16 indicates similarities as well as differences between the roughness and dissonance ratings of the thirteen harmonic intervals, and will be examined in some detail. It is acknowledged that when subjects make dissonance judgments they do not necessarily rely on strictly acoustical criteria. Cultural conditioning has been introduced as an explanation to the consonance / dissonance concept since the 1940s (i.e. Cazden, 1945, in Sethares, 1998: 78-79), while the historical tension between sedimentation and innovation within a musical tradition has supported what Helmholtz recognized as continuously changing attitudes towards consonance and dissonance

(Helmholtz, 1885: 84-85). It may therefore be desirable to include historical and cultural criteria along with physical, physiological, and psychological ones when trying to explain dissonance judgments.

As a first observation, the intervals in Table 6.16 are grouped into eight roughness and eight dissonance categories. However, dissonance categories are more clearly demarcated than roughness categories, indicating that subjects rated dissonance with more confidence than they rated roughness. As opposed to the short history of the concept of roughness, limited within areas outside the mainstream music disciplines, the concept of dissonance has had a long tradition in Western music discourse (Sethares, 1998: 73-80; Hutchinson & Knopoff, 1978: 1; etc.). It should be expected that subjects would be more familiar with the concept of dissonance than with the concept of roughness and would make dissonance ratings with more confidence. The experiment results show signs of categorical perception for dissonance, while roughness ratings seem to have been made along a continuum.

The extremes of the two rating scales in Table 6.16 are occupied by the same intervals (underlined), indicating that the clear presence or absence of roughness dominates dissonance decisions. At the same time, the roughness differences among these harmonic intervals are larger than their dissonance differences. This reduced resolution in the dissonance ratings at the extremes of the scale is consistent with Western culture's preference for smooth sounds over rough ones, offers one more

indication of categorical perception for dissonance, and suggests a possible difference-threshold for the dissonance/annoyance level of roughness.

For intervals located within the extremes of the rating scale, roughness ratings become increasingly ambiguous (increasing overlap among roughness categories) while dissonance ratings become increasingly categorical, demonstrating larger differences than roughness ratings (i.e. the relationship among the Augmented 4th, Fourth, Major 3rd, and Minor 3rd intervals). When the roughness of an interval is neither very large nor very small, dissonance decisions appear to be based on clues additional to roughness, occasionally ignoring roughness altogether (i.e. the relationship between the Augmented 4th and Fourth intervals).

It seems plausible to claim that when there is no clear presence / absence of roughness, dissonance decisions are not based on acoustical cues but are culturally and historically mediated. The high dissonance rating of the Augmented 4th interval (in spite of its lower roughness level than the Major 3rd) and the overall higher dissonance ratings of Minor over Major intervals support such a claim. The Augmented 4th is the only interval within the chromatic scale that is not found in the harmonic series and cannot be arrived at through integer divisions of a string. The mathematical impossibility of the Augmented 4th interval, along with the long-standing link between mathematics and music may have provided the original basis for the unfavorable aura that has followed this interval for centuries.

The clear separation between Major and Minor intervals in the dissonance ratings along with the strong Western-based association of dissonance with unpleasantness reflect a dislike for minor sonorities so ingrained within the Western tradition that it is already present in pre-language children (Kastner & Crowder, 1990). The culture-specific nature of this attitude is indicated by the increase in the distinction between major and minor with age and is fully exposed through encounters with musical traditions where minor harmonies are considered joyful (i.e. music of the Andean highland, Peru. Turino, 1993: 43-58).

The experiment results indicate that roughness constitutes a significant but not the sole factor guiding listeners in their dissonance judgments. Compared to existing theoretical models, the dissonance ranking of the intervals obtained from Experiment 5 agrees best with Stumpf's ranking (1898, in Davies, 1978: 158) except for the Major 7th interval, which was rated less dissonant than predicted by all models other than the proposed model (Eq. (6.23)). Stumpf's dissonance ranking was based on his concept of *fusion*.¹²⁰ The reason for excluding the fusion hypothesis from the present study is that it is based on a large number of interacting and hard to quantify acoustical variables and does not provide a readily measurable physical correlate for dissonance.

¹²⁰ The fusion of two simultaneous tones is proportional to the degree to which the tones are heard as a single perceptual unit. According to Stumpf, fusion is the basis of consonance (Stumpf, 1898, in Sethares, 1998: 77).

For a more detailed examination of the three models addressed in this experiment, Figures 6.28, 6.29, and 6.30 compare the observed roughness and dissonance levels to the roughness values estimated by each model.

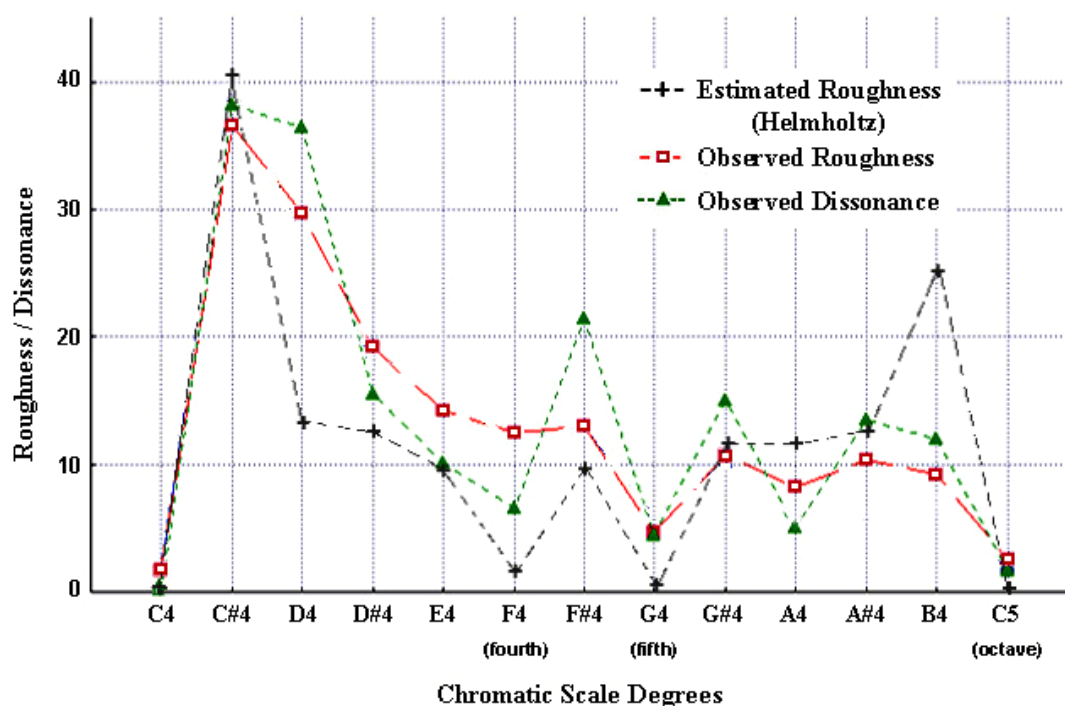


Figure 6.28: Estimated roughness (Helmholtz, 1885) versus observed roughness and observed dissonance for all 13 intervals in the chromatic scale starting at C4 (Experiment 5).

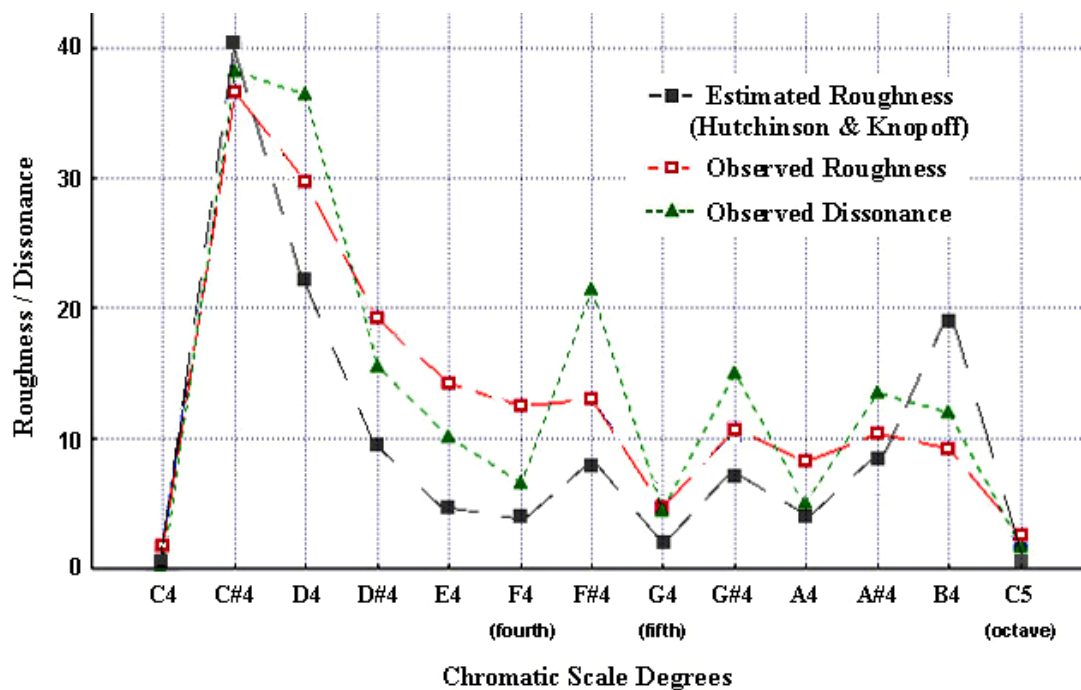


Figure 6.29: Estimated roughness (Hutchinson & Knopoff, 1978) versus observed roughness and observed dissonance for all 13 intervals in the chromatic scale starting at C4 (Experiment 5).

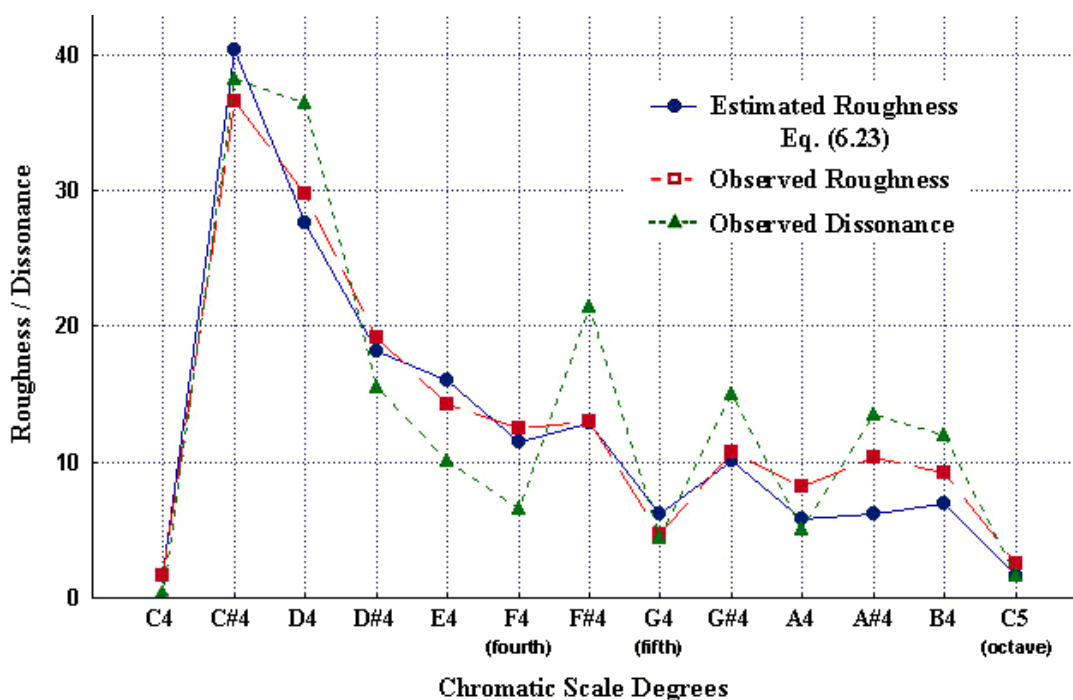


Figure 6.30: Estimated roughness (proposed model, Eq. (6.23)) versus observed roughness and observed dissonance for all 13 intervals in the chromatic scale starting at C4 (Experiment 5).

Table 6.17 displays the Pearson r correlations of estimated roughness to observed roughness and observed dissonance for each of the three roughness estimation models.

Table 6.17 Pearson r Correlations of Estimated Roughness to Observed Roughness and Observed Dissonance for each Roughness Estimation Model		
<u>Roughness Estimation Model</u>	<u>Observed Roughness</u>	<u>Observed Dissonance</u>
Helmholtz, 1885	0.73	0.73
Hutchinson & Knopoff, 1978	0.87	0.87
<u>Proposed Model, Eq. (6.23)</u>	0.98	0.91

The proposed model (Eq. (6.23)) demonstrates the best agreement between estimated and observed roughness ($r = 0.98$) as well as between estimated roughness and observed dissonance ($r = 0.91$). With the exemptions discussed previously, observed dissonance also correlates well with observed roughness ($r = 0.94$).

The results support the claim that, in the Western musical tradition where sensory roughness is avoided as dissonant, the consonance hierarchy of isolated harmonic intervals corresponds mainly to variations in roughness degrees.

Subtle roughness variations are largely ignored. As the presence or absence of roughness become less pronounced, sensory consonance / dissonance gives way to alternative criteria for dissonance judgments. Tenney (1988, in Sethares, 1998: 73-76) discusses five distinct ways in which the consonance and dissonance concept has been understood in music, none of which makes any direct reference to the cultural or the historical dimension of dissonance. The results of the present study suggest that in the absence of strong sensory cues, dissonance judgments of isolated harmonic intervals appear to be culturally and historically mediated.

6.4.4 Summary

Examination of musical instrument construction and performance practice illustrates the musical relevance of amplitude fluctuation and indicates that sound variations involving the sensation of roughness are found in most musical traditions. A new roughness estimation model has been proposed that better represents the theoretical knowledge and experimental results on sensory roughness and is better fit to test hypotheses linking the roughness sensation to musical variables. It was argued that existing models estimating the roughness of pairs of sines demonstrate limited predictive power because they:

- underestimate the contribution of the degree of amplitude fluctuation (i.e. relative amplitudes values of the interfering sines) to roughness,
- overestimate the contribution of sound pressure level (i.e. absolute amplitude values of the interfering sines) to roughness, and
- misrepresent the relationship between roughness and register.

Based on the roughness estimation model introduced by Sethares (1998), a new model has been proposed (Eq. (6.23)), which includes a term to account for the contribution of the amplitudes of interfering sines to the roughness of a sine-pair. This term is based on existing experimental results (von Békésy, 1960; Terhardt, 1974a), adjusted to account for the quantitative difference between amplitude modulation

depth and degree of amplitude fluctuation discussed in Chapter 3. The roughness of complex spectra with more than two sine components is estimated by adding together the roughness of the individual sine-pairs. The proposed model does not account for the influence of phase on roughness (Pressnitzer & McAdams, 1999) and is not fit to handle continuous spectra. To ensure the reliability of the model, the complex spectra examined must: a) have components with the same initial phase *or* less than three components per critical band; b) correspond to signals that are relatively symmetrical along the time axis; and c) be discrete. Under these conditions, the proposed model (Eq. (6.23)) constitutes a good representation of the theoretical knowledge and empirical data on the roughness of complex spectra.

The proposed model was tested experimentally against two earlier roughness estimation models (Helmholtz, 1885; Hutchinson & Knopoff, 1978) and all three models were applied to a hypothesis linking the consonance hierarchy of harmonic intervals, within the Western chromatic scale, to roughness degree variations. The predictions of the three models differed in several respects. For example, the proposed model estimates much lower roughness values than the other models for the Major 7th and the Augmented 4th intervals. The majority of differences in the predictions of the three models can be explained in terms of the different assumptions each model makes regarding the contributions of degree of amplitude fluctuation and of sound pressure level to roughness. The incorrect assumptions of the Helmholtz and

the Hutchinson & Knopoff models result in overestimating the roughness of some sine-pairs and underestimating the roughness of others.

For the experiment, two groups of subjects rated isolated harmonic intervals in terms of roughness and dissonance respectively. The proposed model, Eq. (6.23), demonstrated the best agreement between estimated and observed roughness ($r = 0.98$) as well as between estimated roughness and observed dissonance ($r = 0.91$). The results indicate that the clear presence or absence of roughness in the sound of an interval dominates dissonance ratings. When the roughness is neither very large nor very small, decisions on dissonance often ignore roughness and are culturally and historically mediated. Overall, dissonance ratings correlate well with roughness ratings ($r = 0.94$) indicating that, in the case of isolated harmonic intervals, the sensation of roughness is the primary cue guiding dissonance judgments. The results also support the hypothesis that, in the Western musical tradition where sensory roughness is in general avoided as dissonant, the consonance hierarchy of harmonic intervals corresponds to variations in roughness degrees.

Chapter 7 Summary, Conclusions, and Suggestions for Further Study

7.1 Introduction

The goal of this study has been to improve the understanding of the perceptual and physical properties of amplitude fluctuation and better appreciate its musical significance. Amplitude fluctuation is a manifestation of wave interference and consists of variations in the maximum value of sound signals (i.e. amplitude) as measured relative to a reference point.

The rate, degree, and shape of a complex signal's amplitude fluctuations are variables that are manipulated by musicians around the world to exploit the sensations of beating and roughness, making amplitude fluctuation a significant expressive tool in the production of musical sound. The frequency difference between simultaneous tones influences mainly the rate of fluctuation, the amplitude and phase differences influence mainly the degree of fluctuation, while the number of simultaneous tones and their relative frequency and amplitude values influence the fluctuation shape. The

dissertation reviews critically the history of the understanding of the physical and perceptual properties of waves in general and of amplitude fluctuations in particular, drawing from the areas of music, ethnomusicology, acoustics, psychoacoustics, and engineering. In spite of overwhelming evidence on the energy content of amplitude fluctuation, the related literature contains conflicting opinions regarding its physical properties, as well as inconsistencies in its graphical representation and manipulation. As a result, there are similar inconsistencies regarding the possible perceptual attributes of amplitude fluctuation, hindering the understanding of its musical significance.

The erroneous assumption that signal amplitude fluctuations represent temporal patterns that carry no energy appears to persist mainly because of reports from dichotic experiments. Consequently, the issue of the physical and perceptual properties of amplitude fluctuation has remained unresolved justifying the need for the present study. This Chapter outlines the study findings and discusses possibilities for future research.

7.2 Degree of Amplitude Fluctuation versus Amplitude Modulation Depth

Contrary to convention, degree of amplitude fluctuation (AF-degree) and amplitude modulation depth (AM-depth) are different both conceptually and quantitatively. AF-degree refers to a signal's degree of amplitude fluctuation, while AM-depth refers to a signal's degree of spectral energy spread. Whenever AM-depth values are used as a measure of a modulated signal's AF-degree, modulation implementation produces an error (Eq. (3.6)) arising from the nonlinear relationship between presumed and applied AF-degrees.

It was shown that, in order to apply an intended AF-degree, an adjusted coefficient m has to be inserted in the modulation implementation equation, Eq. (3.1). This necessary adjustment (Eq. (3.8)) influences the interpretation of studies where presumed changes in the degree of amplitude fluctuation have been correlated to changes in physical, physiological or psychological measurements. Consequently, studies that manipulate AF-degree through systematic implementation of AM-depth may need to be revised.

In a study by Terhardt (1974a), for example, the identified error provided unjustified support to a hypothesis of a neural low-pass weighting mechanism operating on a signal's temporal envelope. Additionally, it has supported roughness estimation models of complex spectra that largely underestimate the influence of the degree of amplitude fluctuation to the roughness of complex tones.

In another example, the psychometric functions provided by Ritsma (1962, 1963) have underestimated the importance of amplitude fluctuation degree to the salience of the residue pitch of AM tones. The proposed adjustment suggests that reducing the degree of amplitude fluctuation results in a drastic reduction in the existence region of the tonal residue (i.e. frequency range within which higher components can give rise to the perception of the residue pitch). This change in interpretation has implications to pitch perception models that relied on the originally reported results.

As a general observation, the erroneous assumption of the equivalence between amplitude modulation and amplitude fluctuation has resulted in the misinterpretation of studies examining perceptual correlates of amplitude fluctuation, providing questionable support to related perceptual models. Psychometric functions from studies that have miscalculated the degree of amplitude fluctuation have been distorted by a factor proportional to the error (Eq. (3.6)). The influence of the proposed adjustment to the interpretation of distorted psychometric functions will vary according to the meaning attributed by each study to each specific function's shape.

Careful reexamination of such studies is necessary, especially if their results constitute the principal support to a larger model.

The present dissertation kept the terms *amplitude modulation* and *amplitude fluctuation* separate and adopted the term *amplitude fluctuation* to refer specifically to variations of a signal's amplitude around a reference value.

7.3 Theoretical and Experimental Examination of the Wave Energy Carried by Amplitude Fluctuations

Sound wave interference gives rise to amplitude fluctuations, indicating that the perception of sound interference products may be a manifestation of amplitude fluctuation energy. It is argued that modulations in general and amplitude fluctuations in particular satisfy all criteria that define a wave, including the behavior expected from waves in the case of dispersion. They propagate with the *group velocity* v_{gr} , distinguished from the propagation velocity of the carriers, the *phase velocity* v_{ph} . Group and phase velocities describe two distinct physical variables that, in the case of non-dispersive media such as air, simply share the same value. Although air is a non-dispersive medium, the dispersive case is addressed because it facilitates the analysis

of the kinematic properties of modulated waves. Introducing the concept of dispersion to the study of musical sound is not unusual. Sound diffraction, absorption, and reverberation have all been better understood with the aid of dispersion theory.

The energy content of amplitude fluctuations was investigated in an effort to provide theoretical and experimental support to the large amount of evidence demonstrating energy transfer at the amplitude fluctuation rate:

- Amplitude fluctuations can transfer energy to systems (i.e. machinery, resonators, etc.) whose natural frequency matches the fluctuation rate.
- Amplitude fluctuations of acoustic waves correspond to changes in the vibration velocity of a mass at a rate equal to the rate of fluctuation, indicating consumption / radiation of energy at the same rate.
- Amplitude fluctuations entering a dispersive medium separate from the wave that carries them and travel independently at an observable rate and strength.
- Successive portions of a diffracted wave-front are determined by the envelope of secondary wavelets, indicating that amplitude fluctuations and the wave envelopes that describe them behave as waves.

A theoretical value for the energy content of amplitude fluctuations was calculated for the simplified case of a single point-source oscillating at two angular frequencies ω_1, ω_2 , with the same amplitudes $A_1 = A_2 = A$, and zero initial phases:

$$y(t) = A \cos(\omega_1 t) + A \cos(\omega_2 t) = 2A \cos \omega_{\text{mod}} t (\cos \omega_{\text{av}} t) \quad \text{Eq. (6.2)}$$

$$\text{where} \quad \omega_{\text{av}} = \frac{\omega_2 + \omega_1}{2} \gg \omega_{\text{mod}} = \frac{|\omega_2 - \omega_1|}{2}.$$

Eq. (6.2) describes a two-component complex wave as a sine wave with fluctuating amplitude. Assuming that the conditions permitting linear addition of wave energy quantities are satisfied, the application of the energy conservation law indicates a power difference between the two sides of Eq. (6.2). This difference is interpreted as the power radiated by the fluctuations P_{AF} .

$$P_{AF} = \frac{\omega_{AF}^2 \rho A_{AF}^2}{8\pi c} \quad \text{Eq. (6.6)}$$

Eq (6.6) describes the power radiated by a point-source pulsating with angular frequency $\omega_{AF} = |\omega_1 - \omega_2|$ (rate of amplitude fluctuation) and amplitude $A_{AF} = \sqrt{0.5}A = 0.707A$, where ω_1 , ω_2 , and A are the angular frequencies and common amplitude of the interfering sines respectively.

To empirically verify the energy content of amplitude fluctuations, the predicted frequency and amplitude values were compared to values obtained through Hilbert demodulation of the complex signal in Eq. (6.2) and Frequency Zoom analysis of the resulting envelope signal. The analysis revealed a single component with

frequency equal to the complex signal's amplitude fluctuation rate, providing clear evidence of the energy content of amplitude fluctuations. However, the amplitude value of the fluctuation component indicated by the analysis, $A_{AF} = 0.776A$, was different from the predicted value. The discrepancy between predicted and observed amplitudes may be attributed to possible violations of the assumptions underlying the application of the energy conservation law, as well as to inadequate protection from analysis errors due to leakage. The experiment confirms the energy content of amplitude fluctuations but is quantitatively inconclusive, indicating the need for further study.

The theoretical model proposed to predict the power radiated at the rate of amplitude fluctuation is limited to the simplest possible case and relies on energy conservation assumptions that further restrict its use. A more suitable model would be based on a valid modification / simplification of the generalized expressions for the pressure at the difference frequency (in Hamilton, 1998: 223-225). Additionally, the methods employed to empirically verify the predictions of the proposed model are indirect and prone to analysis errors. A direct experimental method would employ non-intrusive measurement techniques (i.e. laser vibrometry), applied on finely tuned resonators exposed to the experiment signal and placed within an environment isolated from external vibrations. It is expected that such a method will verify the energy content of amplitude fluctuations, putting an end to two hundred years of controversy and informing future research in music perception and cognition.

7.4 Alternative Representation of Sound Waves for Better Illustration of Wave Energy Quantities

An alternative sound signal representation has been proposed that a) represents graphically the energy content of amplitude fluctuations predicted by Eq. (6.6) and b) solves the negative-amplitude and consistent-energy-representation problems associated with traditional two-dimensional signals. The new signal representation is based on the complex equation of motion and includes both the sine (imaginary) and cosine (real) terms describing a vibration. It results in spiral sine signals and twisted-spiral complex signals (Figures 4.1, 6.13, & 6.14 and the Appendix), similar to complex analytic signals. Three-dimensional signals illustrate that amplitude fluctuations and the signal envelopes that describe them are not just boundary curves but waves that trace changes in the total instantaneous energy of a signal over time, representing the oscillation between potential and kinetic energies within a wave.

Three-dimensional sound signals are not intended to replace their two-dimensional counterparts but to inform them. Spiral sine signals offer a consistent measure of the energy content of sine waves across frequencies, while twisted spiral complex signals account for the negative amplitude values predicted by the

mathematical expression of amplitude fluctuation, map the parameters of amplitude fluctuation onto the twisting parameters, and paint a more realistic picture of wave propagation. Developing software that displays in real time an arbitrary, user-defined complex wave in three dimensions as a twisted spiral will contribute to the better understanding of sound wave propagation in air.

7.5 Sound Interference Products and Neural Wave Interaction

The fact that amplitude fluctuations carry energy has been challenged by previous experiments where sound interference products appear to arise at a neural level. The possibility of interference products arising from neural wave interaction was examined through a series of dichotic experiments. It was shown that the perception of sound interference in terms of loudness changes, beating, or roughness does not arise unless sound waves interact physically.

Three sets of dichotic experiments confirmed this claim. The first experiment examined the possibility of beating when a sine stimulus is presented to both ears through headphones. In this case, diplacusis causes a pitch difference between the two

ears that would introduce beating if neural wave interaction could give rise to interference products. However, no such beating was observed.

The second set of experiments examined the effect of phase on dichotically presented sine stimuli. Results indicated that changes in the phase relationship between the left and right channels of dichotic sine signals give rise to perceptions that are incompatible with the interference principle and consistent with results from studies that have established phase differences between ears as sound localization cues. It was shown that:

- Changing the phase relationship between the left and right channels of a sinusoidal stimulus presented dichotically does not change its loudness.
- For frequencies <500Hz (200Hz & 300Hz), dichotic phase changes affect the perceived direction of a stimulus, with 180^0 phase difference resulting in a wider stereo image.
- For frequencies >1500Hz (1700Hz & 2500Hz), dichotic phase changes introduce no perceptual change.

The results support the claim that the beating-like sensation reported in previous dichotic experiments is not an indication of interference products arising at a neural level but the manifestation of a sound localization mechanism based on phase difference cues.

The last experiment was designed to examine this claim. It was shown that, in dichotic explorations of beats, the constantly shifting phase between the sine components presented in the two ears results in a constantly shifting localization of the two-component complex stimulus. For low frequency differences ($\approx \leq 3\text{Hz}$) this shift is interpreted as a rotation of the sound inside the subject's head and, for higher frequency differences ($\approx \geq 3\text{Hz}$) it is interpreted as a timbral fluctuation that is often confused by the subjects with loudness fluctuation or beating. This confusion was shown to always arise when comparing rotating and modulated sines with rotation / modulation rates $\approx \geq 3\text{Hz}$. Consistent with sound localization studies and contrary to the interference principle, it was shown that two-component dichotic stimuli with different frequencies in each ear are treated perceptually as rotating sines for frequencies $< 500\text{Hz}$ (300Hz). For frequencies $> 1500\text{Hz}$ (1700Hz) they are treated as steady sines, with the frequency difference between ears introducing no perceptual difference.

The findings support the hypothesis against sound interference products arising due to neural wave interaction. They are consistent with the arguments for the energy content of amplitude fluctuation and indicate that the beating sensation reported in previous dichotic experiments must have been a misidentified rotating sensation.

7.6 Auditory Roughness Estimation of Complex Spectra - Roughness Degrees and Harmonic Interval Dissonance Ratings within the Western Musical Tradition

Examination of musical instrument construction and performance practice illustrates the musical relevance of amplitude fluctuation energy and indicates that sound variations involving the sensation of roughness are present in most musical traditions. A new roughness estimation model has been proposed that better represents the theoretical knowledge and experimental results on sensory roughness and is better fit to test hypotheses linking the roughness sensation to musical variables. It was argued that existing models estimating the roughness of sine-pairs demonstrate limited predictive power because they:

- underestimate the contribution of the degree of amplitude fluctuation (i.e. relative amplitudes values of the interfering sines) to roughness,
- overestimate the contribution of sound pressure level (i.e. absolute amplitude values of the interfering sines) to roughness, and
- misrepresent the relationship between roughness and register.

Based on the roughness estimation model introduced by Sethares (1998), a new model was proposed (Eq. (6.23)), which includes a term to account for the contribution of the amplitudes of interfering sines to the roughness of a sine-pair. This term is based on existing experimental results (von Békésy, 1960; Terhardt, 1974a), adjusted to account for the quantitative difference between amplitude modulation depth and degree of amplitude fluctuation identified by the present study. The roughness of complex spectra with more than two sine components is estimated by adding together the roughness of the individual sine-pairs. The proposed model does not account for the influence of phase on roughness and is not fit to handle continuous spectra. To ensure the reliability of the model, the complex spectra examined must: a) have components with the same initial phase *or* less than three components per critical band; b) correspond to signals that are relatively symmetrical along the time axis; and c) be discrete. Under these conditions, the proposed model (Eq. (6.23)) constitutes a good representation of the theoretical knowledge and empirical data on the roughness of complex spectra.

The proposed model was tested experimentally against two earlier roughness estimation models (Helmholtz, 1885; Hutchinson & Knopoff, 1978), and all three models were applied to a hypothesis linking the consonance hierarchy of harmonic intervals, within the Western chromatic scale, to roughness variations. The predictions of the three models differ in several respects. For example, the proposed model estimates much lower roughness values than the other models for the Major 7th

and the Augmented 4th intervals. The majority of differences in the predictions of the three models can be explained in terms of the different assumptions each model makes with regards to the contributions of degree of amplitude fluctuation and of sound pressure level to roughness. The incorrect assumptions of the Helmholtz and the Hutchinson & Knopoff models result in overestimating the roughness of some sine-pairs and underestimating the roughness of others. The proposed model, Eq. (6.23), demonstrates the best agreement between estimated and observed roughness as well as between estimated roughness and observed dissonance.

The empirical results indicate that the clear presence or absence of roughness in the sound of an interval dominates dissonance ratings. When the roughness is neither very large nor very small, decisions on dissonance often ignore roughness and are culturally and historically mediated. Overall, dissonance ratings correlate well with roughness ratings, supporting the claim that, in the case of isolated harmonic intervals, the sensation of roughness is the primary cue guiding dissonance judgments. The results also support the hypothesis that, in the Western musical tradition where sensory roughness is in general avoided as dissonant, the consonance hierarchy of harmonic intervals corresponds to variations in roughness degrees.

The validity and utility of the proposed roughness estimation model can be improved significantly through a series of theoretical and practical modifications. The

following steps would remove the limiting conditions imposed on the model with regards to the effects of phase and of signal envelope asymmetry on roughness:

- Two functions must be introduced. One to describe the relationship between the relative phase of three interfering sines and the degree of amplitude fluctuation of the resulting three-component signal, and a second one to quantify signal envelope asymmetry.
- Two sets of perceptual experiment must be conducted. One to examine the roughness relationship between AM tones with a specified AF-degree and AM tones with the same AF-degree arrived at through phase manipulation and a second one to examine the effect of signal envelope asymmetry on roughness.
- Two coefficients must be added to the proposed roughness estimation model (Eq. (6.23)), implementing the results from the previous steps.

The modified model will represent a comprehensive account of the current knowledge on sensory roughness and will be valid to the examination of all discrete spectra.

To improve the utility of the proposed model and make it available to a great variety of research questions a computer application must be created that will:

- perform spectral analysis on incoming sound signals at user-specified time intervals and measure the signal's degree of asymmetry based on the previously identified asymmetry function;

- process the spectral analysis results based on an algorithm that selects the significant (according to user-specified criteria) components of the spectrum and output a frequency, an amplitude, and a phase value for each selected component;
- supply the frequency, amplitude, phase, and asymmetry values to the roughness estimation model;
- provide numerical and graphical versions of the model's output.

A computer application such as the one described would enable the empirical testing of hypotheses linking roughness to concepts of musical tension, release, stability, etc., which are temporal and cannot be addressed efficiently by calculating the roughness of isolated sonorities. Additionally, automating the roughness estimation process would make the model easy to apply, guaranteeing its use in future research.

7.7 Manipulation of Amplitude Fluctuation and Musical Expression

Amplitude fluctuation constitutes the most common time-variant characteristic of musical sounds. The large variety of perceptual possibilities introduced through the manipulation of amplitude fluctuation (i.e. beating, roughness, combination tones) has been exploited by numerous musical traditions throughout history. Manipulating the degree and rate of amplitude fluctuation helps create a shimmering (i.e. Indonesian *gamelan* performances), buzzing (i.e. Indian *tambura* drone), or rattling (i.e. Bosnian *ganga* singing) sonic canvas that becomes the backdrop for further musical elaboration. It permits the creation of timbral (i.e. *mijwiz* playing) or even rhythmic (i.e. *ganga* singing) variations through gradual or abrupt changes between fluctuation rates and degrees. Whether such variations are explicitly sought for (as in *ganga* singing and *mijwiz* playing) or are introduced more subtly and gradually (as may be the case in the typical chord progressions / modulations of Western music), they have both sonic and symbolic significance and form an important part of a musical tradition's expressive vocabulary.

Important clues regarding the ways various musical cultures approach the perceptual attributes of amplitude fluctuation were identified through an examination

of musical instrument construction and performance practice. For example, the Middle Eastern *mijwiz* owes its rich tonal quality to the slight detuning between its two, otherwise identical, cane pipes, giving tones that beat constantly and at slightly shifting rates. Desired variations among roughness degrees are achieved by occasionally increasing the detuning of the near-Unisons through partial stopping or by temporarily abandoning the Unisons and performing a lower melody on one pipe over a high drone on the other. In its construction and performance practice, the *mijwiz* is an example of an instrument that makes explicit use of the perceptual richness of amplitude fluctuation through creative exploration of the beating and roughness sensations. Along with the rest of the examples cited in this dissertation, it helps illustrate the significance of a study on amplitude fluctuation to music.

Explicit understanding of the physical variables associated with the various aspects of musical sound is not a prerequisite to music making or music appreciation. Music has been produced, listened to, and appreciated long before a good understanding of the physical properties of sound had been reached. Such an understanding, however, is valuable in helping explain and understand diverse musical choices and performance practices, as well as in offering the opportunity to explore new sonic effects through the informed manipulation of physical variables.

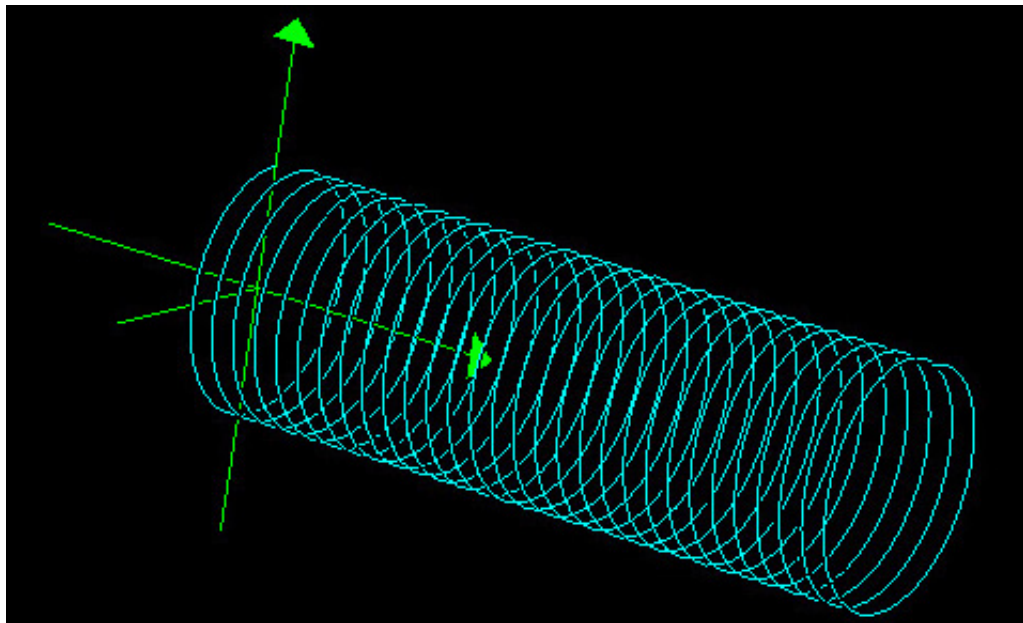
A Western musician may justify the use of certain harmonies based on the concepts of consonance / dissonance, while a Bosnian singer may claim that narrow intervals are used partly because they sound louder than Unisons or wide intervals.

Better understanding of the perceptual and physical properties of amplitude fluctuation may lead to a better understanding of the basis of such claims. Additionally, whenever a Bosnian listener judges a Western piece of music as 'blunt' or a Western listener calls a *ganga* song 'offensive', perceptual differentiation has preceded evaluative judgment.

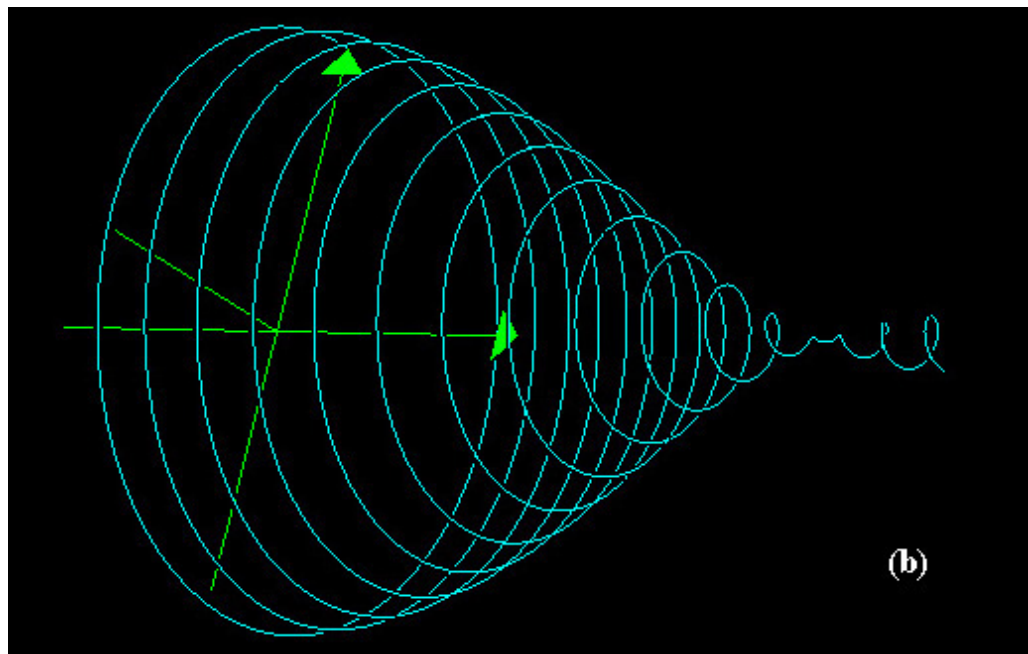
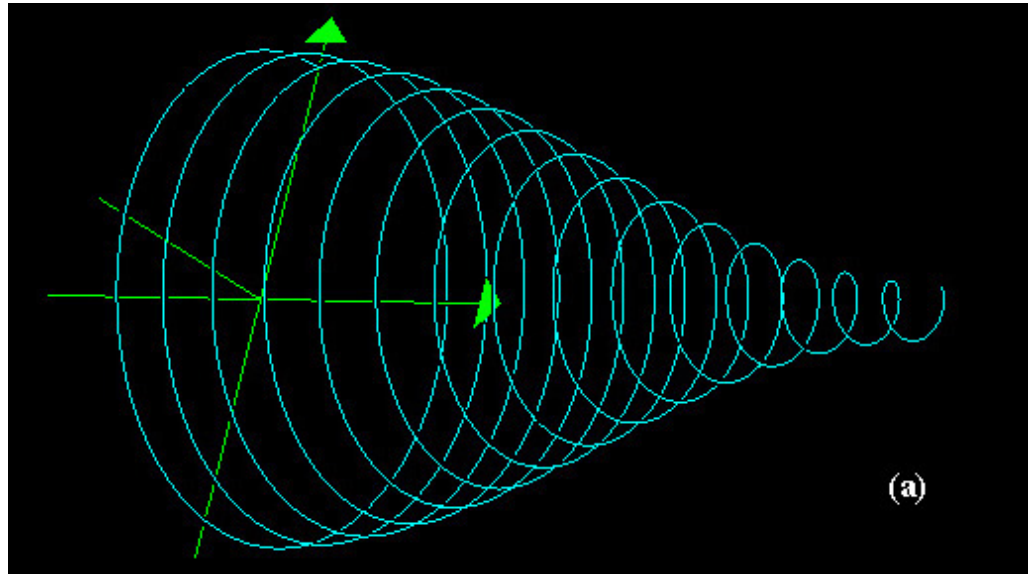
Regardless of whether more or less beating / roughness is called for within a musical context, or whether a specific relationship between interfering tones and difference tones is more desirable than others, being able to control their production allows for the creation of finer sonic nuances. More importantly, it helps in appreciating that musical evaluations are based on cultural ideals and do not map on levels of intellectual sophistication. All sounds, regardless of whether they are evaluated as 'good' or 'bad', result from the manipulation of the same variables. The production of a musical sound that is considered 'bad', within a musical tradition, may reflect a more sophisticated manipulation of certain sonic variables than the production of a sound that is considered 'good'. The present study improved significantly the understanding of amplitude fluctuation by examining musical practices considered, until recently, inferior. It is hoped that this work will motivate the growth of cross-cultural, interdisciplinary research in musical sound and offer helpful tools for the analysis, understanding, appreciation, and production of music.

Appendix

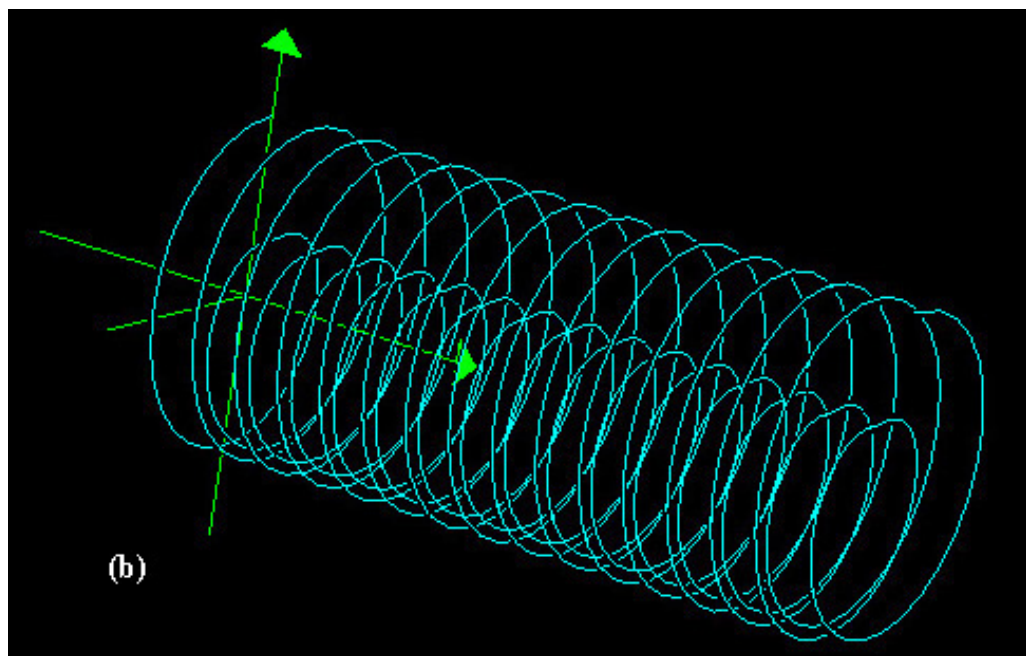
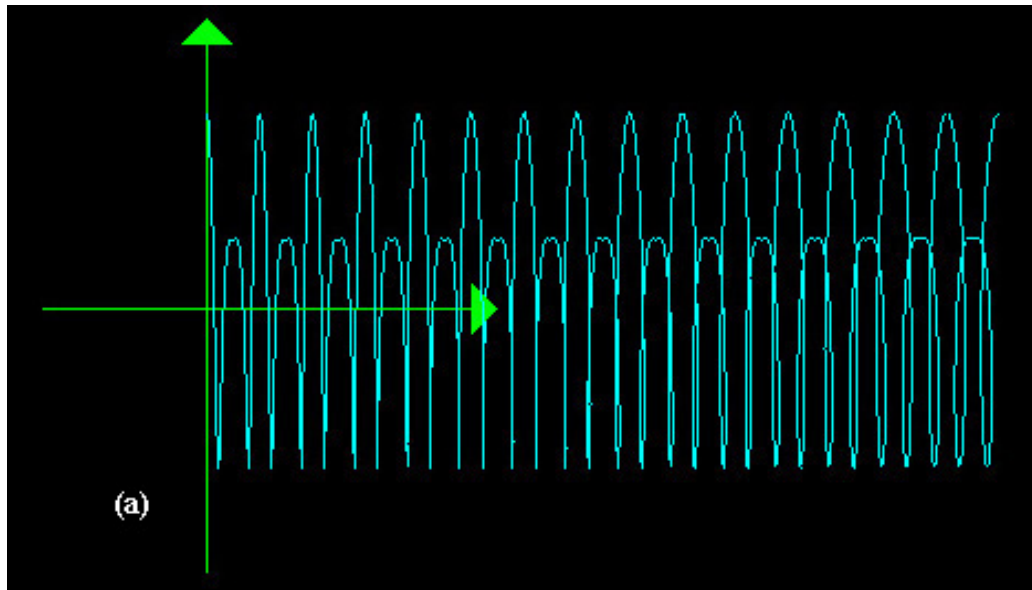
The present Section displays examples of spiral sine signals and twisted spiral complex signals, designed using *NuCalc 2.0* by Pacific Tech (1998). The examples outline the relationship between two- and three-dimensional signals and illustrate their relevance to the study of amplitude fluctuation. More details in Sections 5.2.1 & 6.3.2.



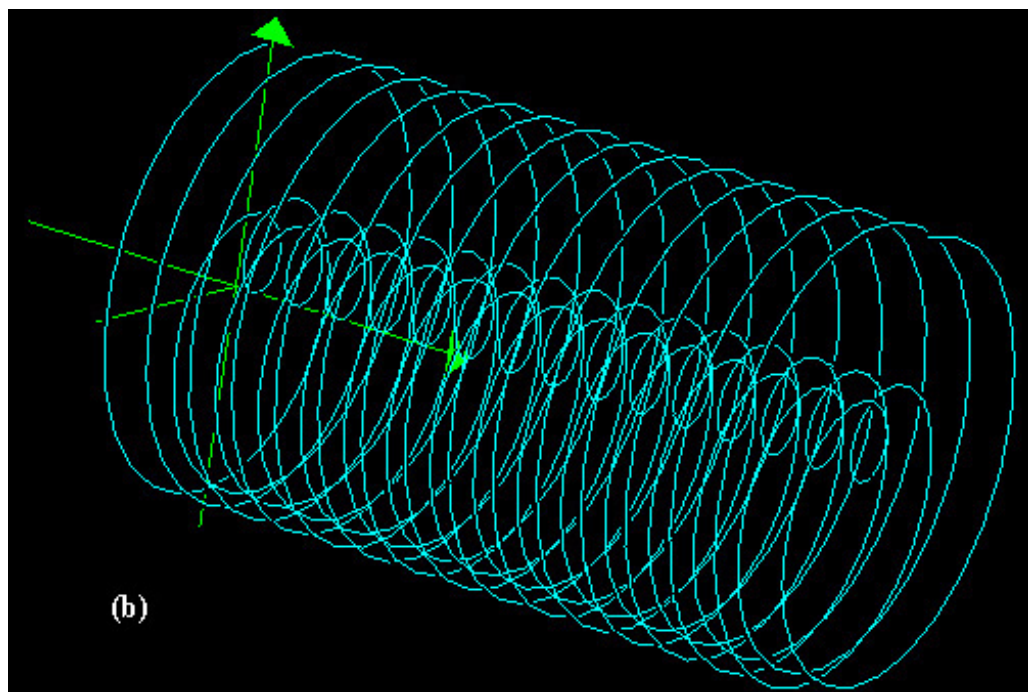
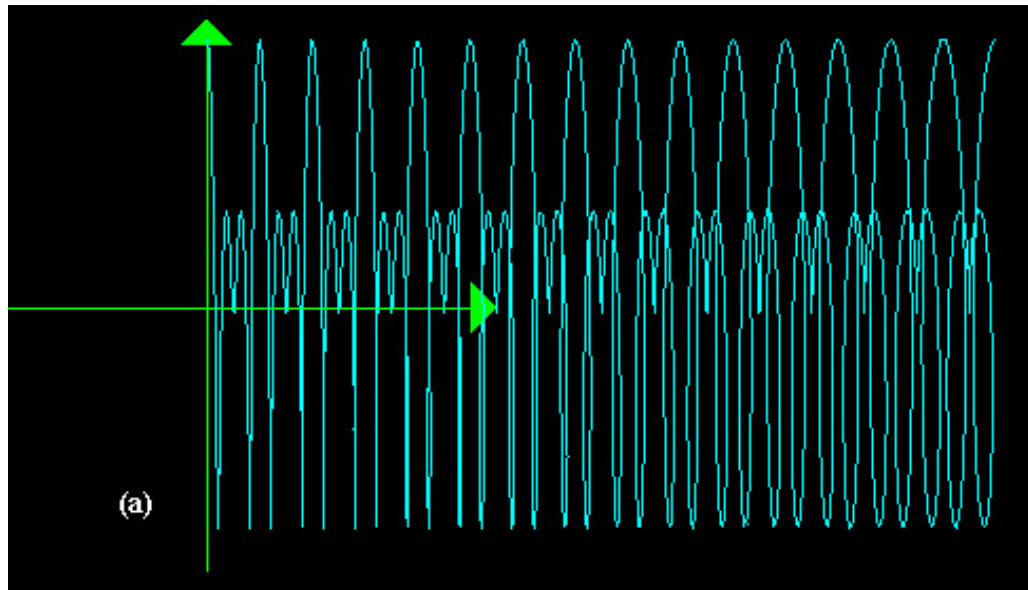
Three-dimensional spiral cosine signal with frequency $f = 30\text{Hz}$ and duration $t = 1\text{sec}$. Horizontal arrow: time; vertical arrow: cosine (real axis); third axis: sine (imaginary axis). In three-dimensional spiral signals, the power of sinusoidal waves (total area outlined by all cycles) is represented consistently across frequencies.



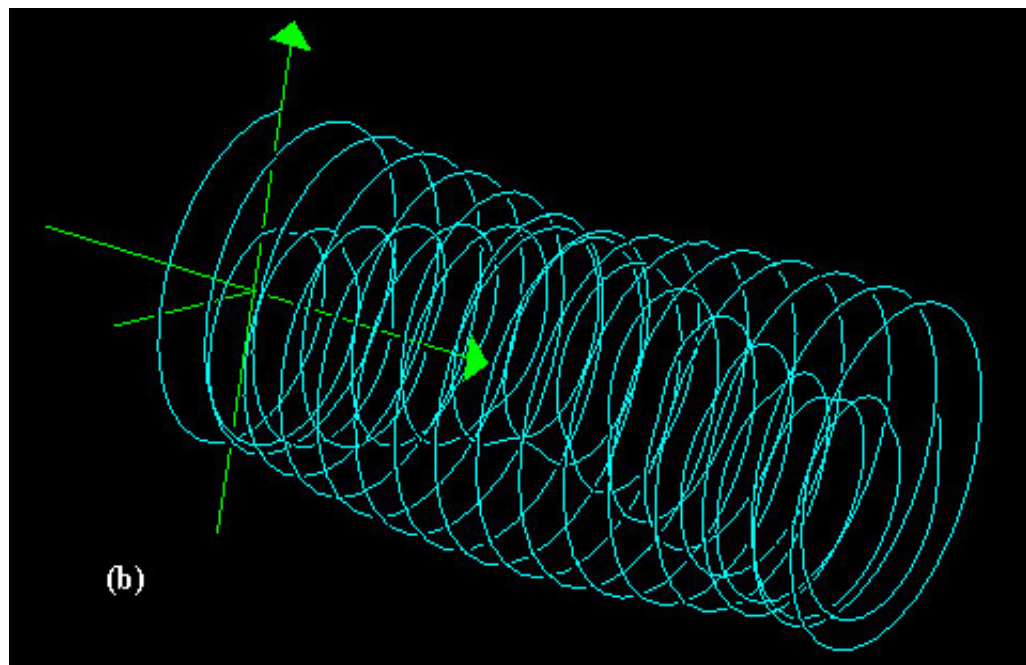
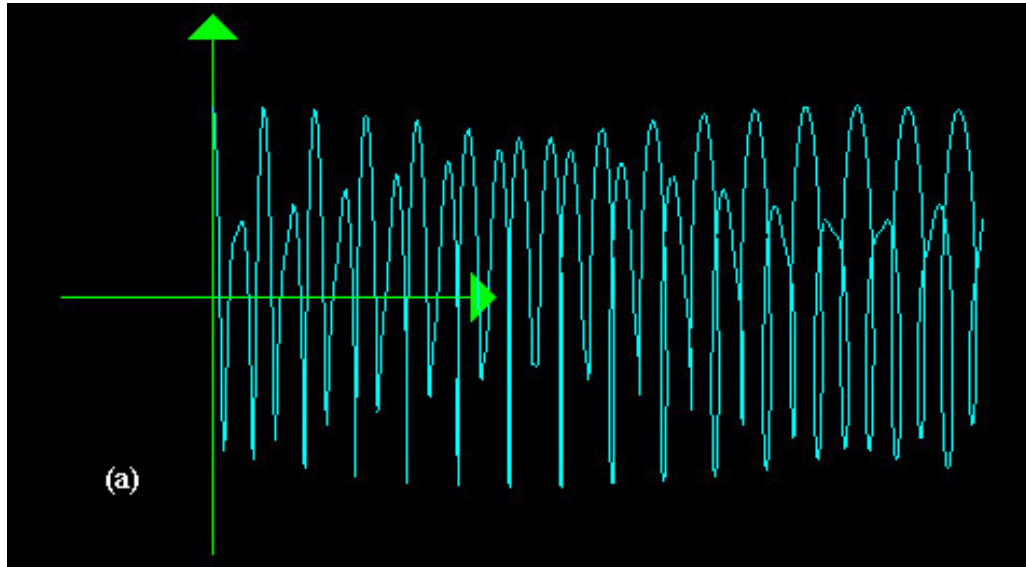
First half-second of a 30Hz cosine, amplitude modulated with $f_{mod} = 1\text{Hz}$ and (a) $AM_{depth} = 0.8$ or (b) $AM_{depth} = 1.2$. When AM_{depth} increases beyond 1 (100%), the signal is twisted around the horizontal axis, representing the alternating positive and negative amplitude values of AM signals with $AM_{depth} > 1$.



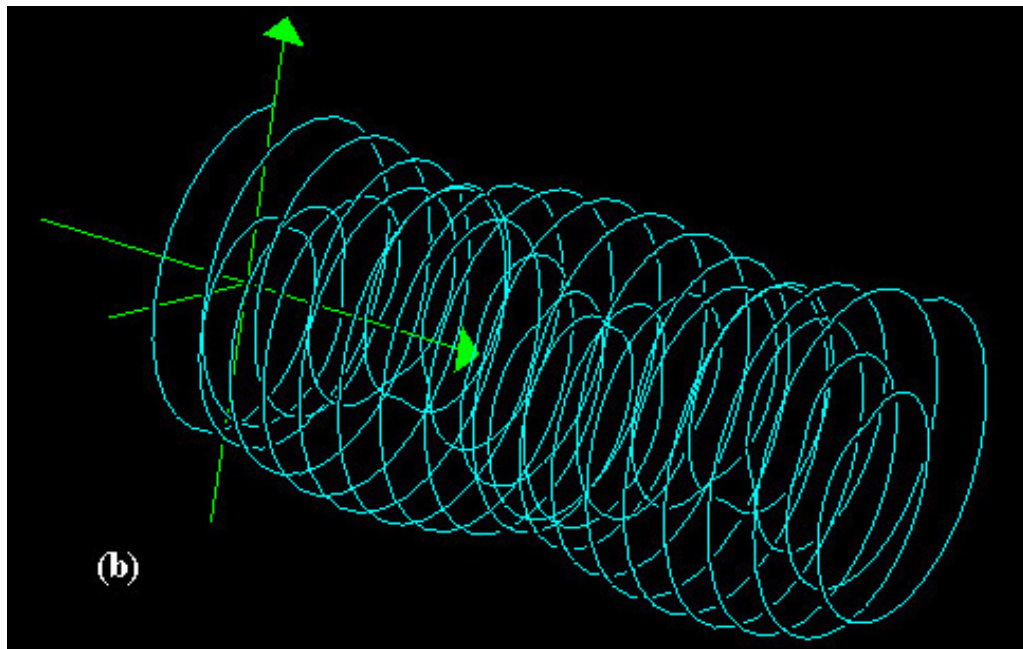
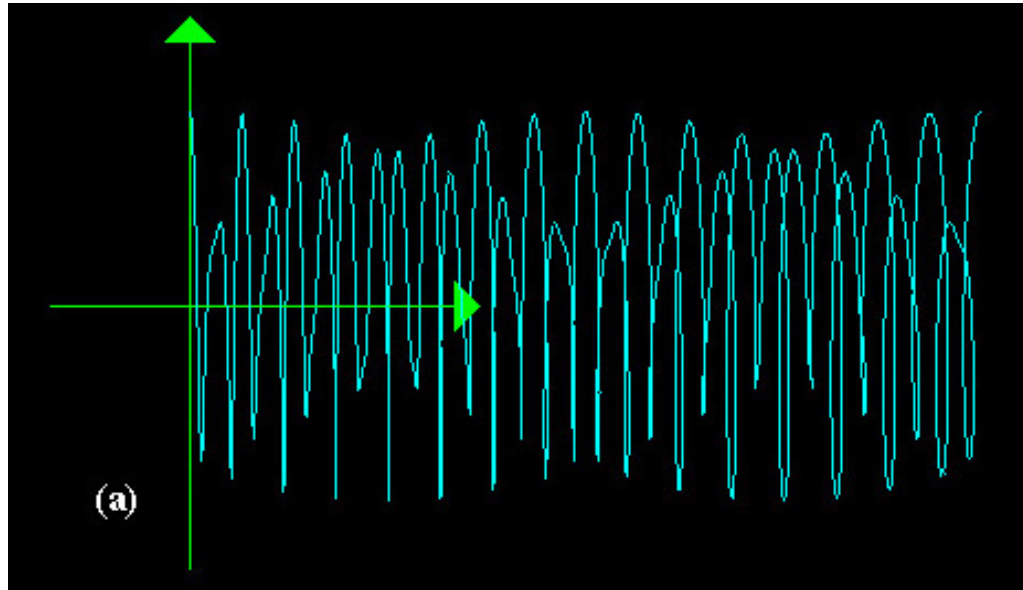
(a) Two- and (b) three-dimensional signal of a Fifth interval (just intonation): $f_1 = 30\text{Hz}$; $f_2 = 45\text{Hz}$; $A_1 = 1$; $A_2 = 0.5$. The two-dimensional signal in (a) derives from projecting the three-dimensional signal onto two dimensions consisting of the cosine axis (vertical arrow) and the time axis (horizontal arrow).



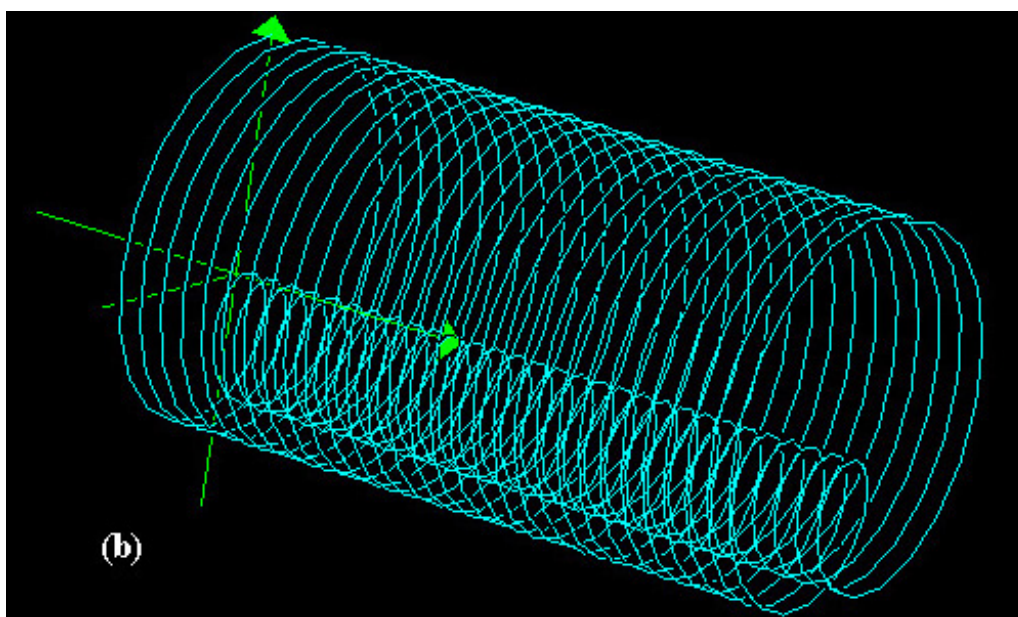
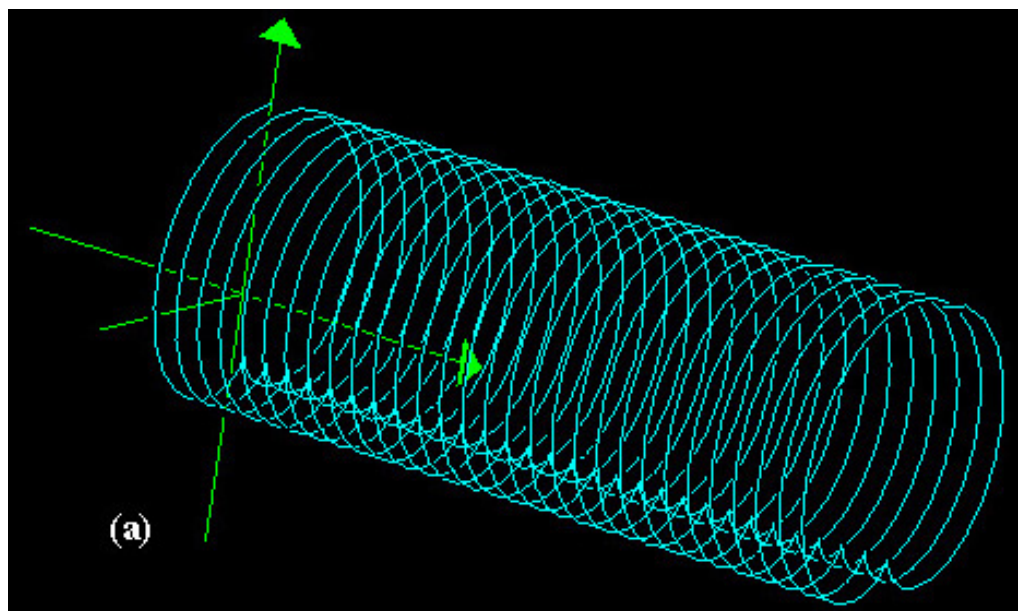
Again, (a) two- and (b) three-dimensional signal of a Fifth: $f_1 = 30\text{Hz}$; $f_2 = 45\text{Hz}$. Now, however, $A_1 = A_2 = 1$. As the amplitude of the higher frequency tone increases from $A_2 = 0.5$ to $A_2 = 1$, the signal twists to include the additional 15 cycles.



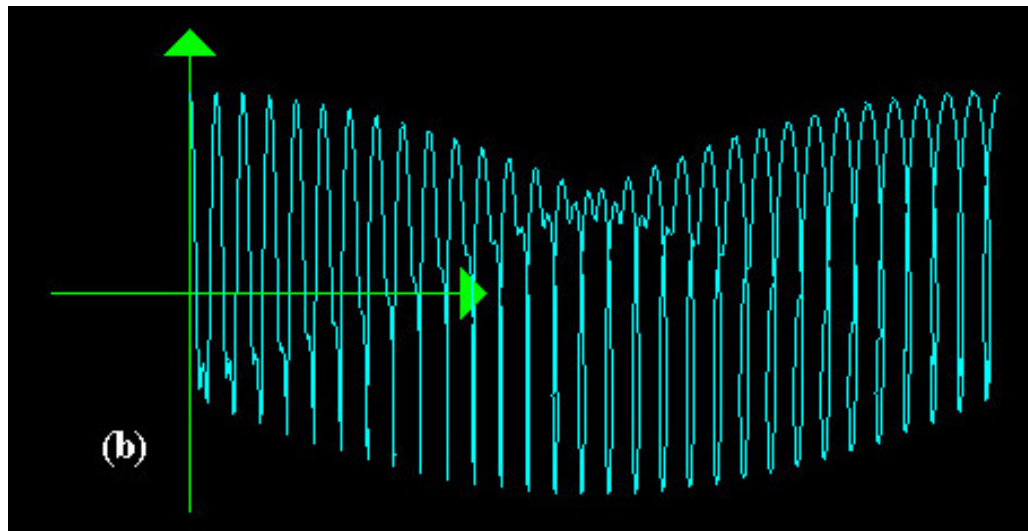
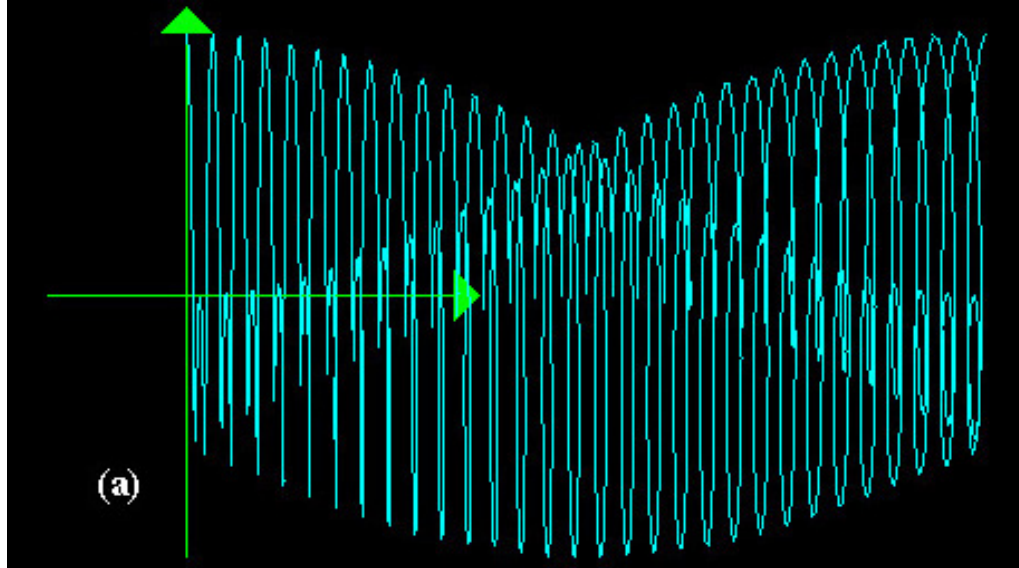
(a) Two- and (b) three-dimensional signal of a mistuned Fifth: $f_1 = 30\text{Hz}$; $f_2 = 45\text{Hz} + d(0.6\text{Hz}) = 45.6\text{Hz}$; $A_1 = 1$; $A_2 = 0.5$. The beating rate is equal to $2df = 1.2\text{Hz}$, and equal to the ‘bending’ rate of the two intertwined spirals (combined amplitude fluctuation rate of the twisted spiral).



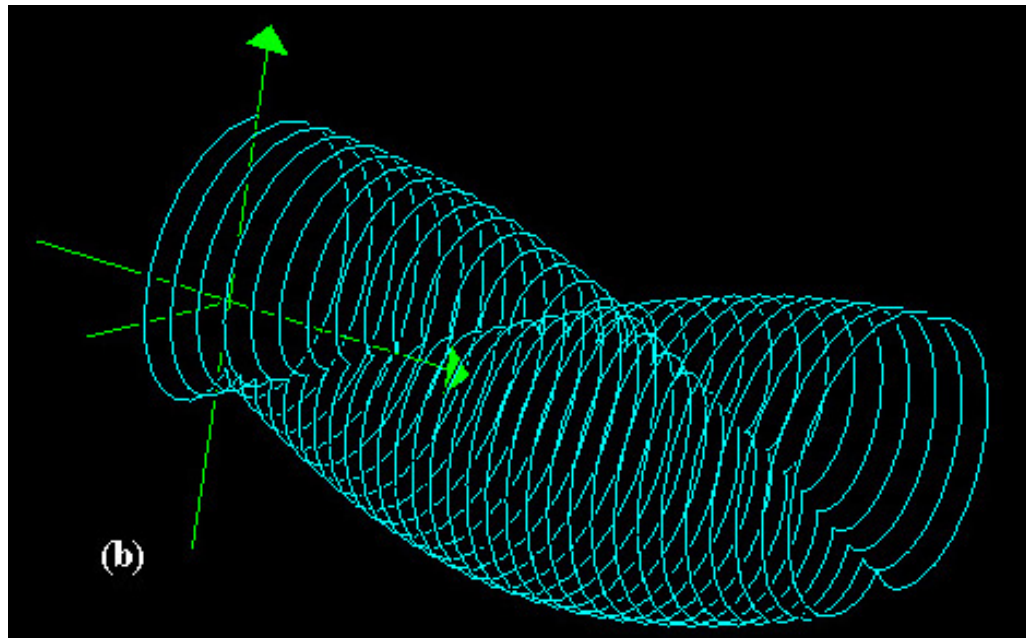
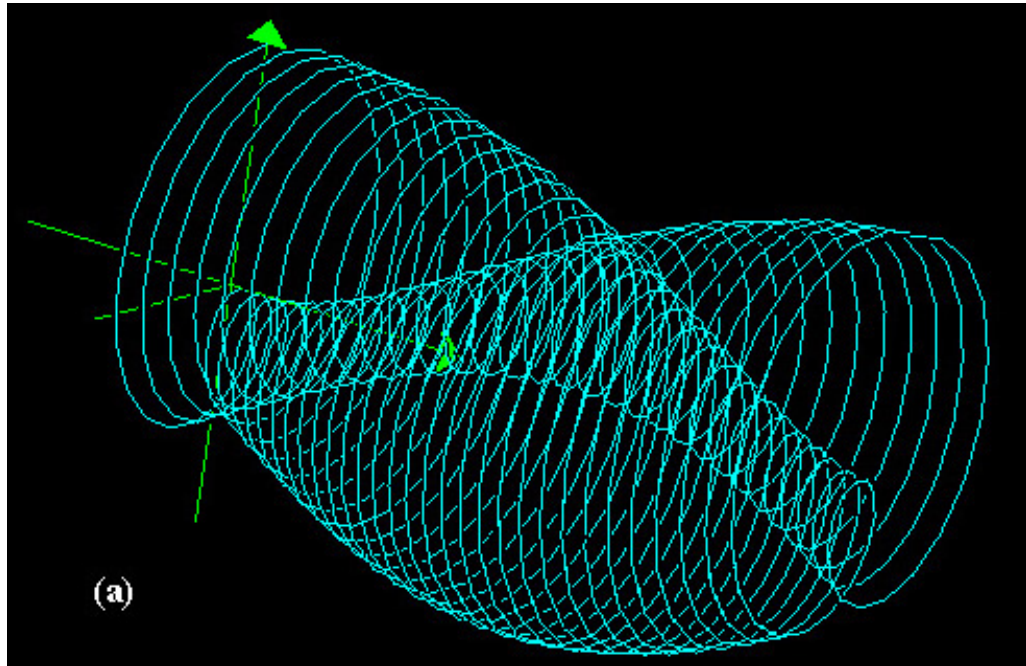
Again (a) two- and (b) three-dimensional signal of a mistuned Fifth. In this case, $f_1 = 30\text{Hz}$; $f_2 = 45\text{Hz} + df(1\text{Hz}) = 46\text{Hz}$; $A_1 = 1$; $A_2 = 0.5$. The beating rate is again equal to the ‘bending’ rate of the two intertwined spirals ($2df = 2\text{Hz}$; combined amplitude fluctuation rate of the twisted spiral).



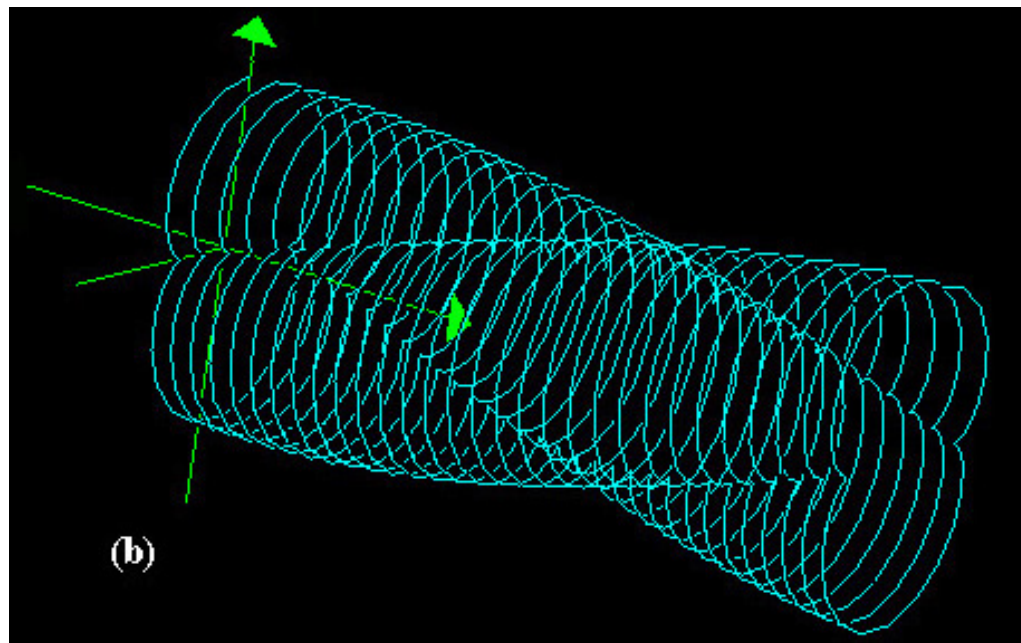
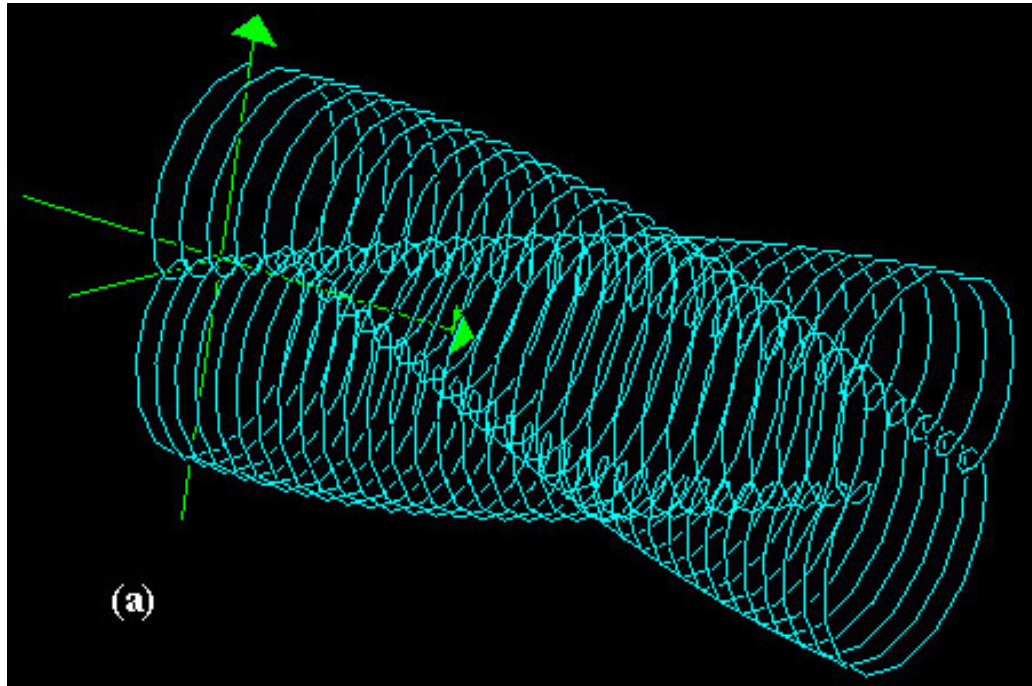
Three-dimensional signals of an Octave: $f_1 = 30\text{Hz}$; $f_2 = 60\text{Hz}$; $A_1 = 1$; and
 (a) $A_2 = 0.5$ or (b) $A_2 = 1$. As was the case with the Fifth, when the amplitude of the
 higher frequency tone increases from $A_2 = 0.5$ to $A_2 = 1$, the signal twists to include the
 additional cycles.



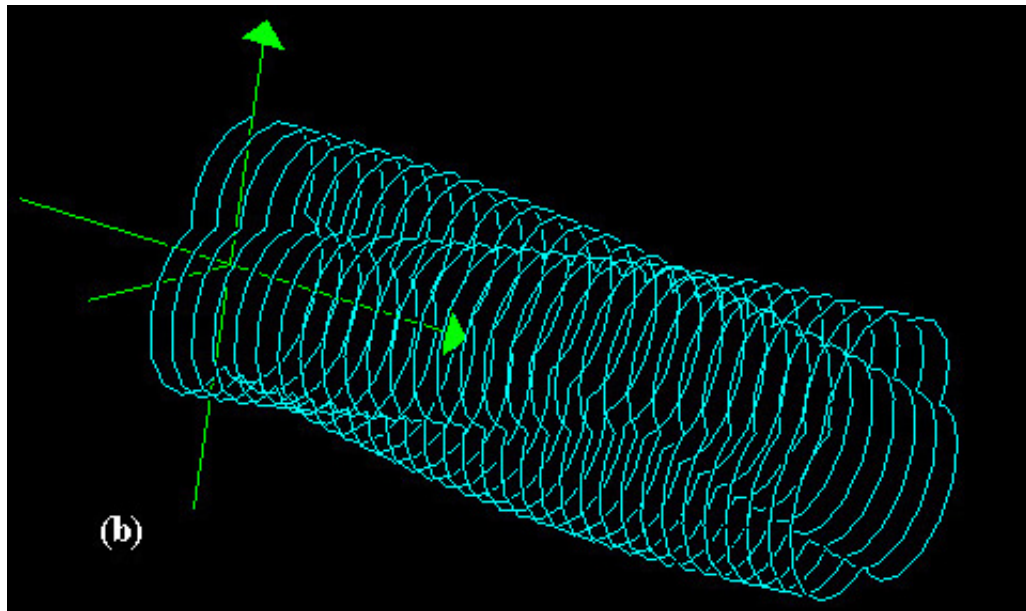
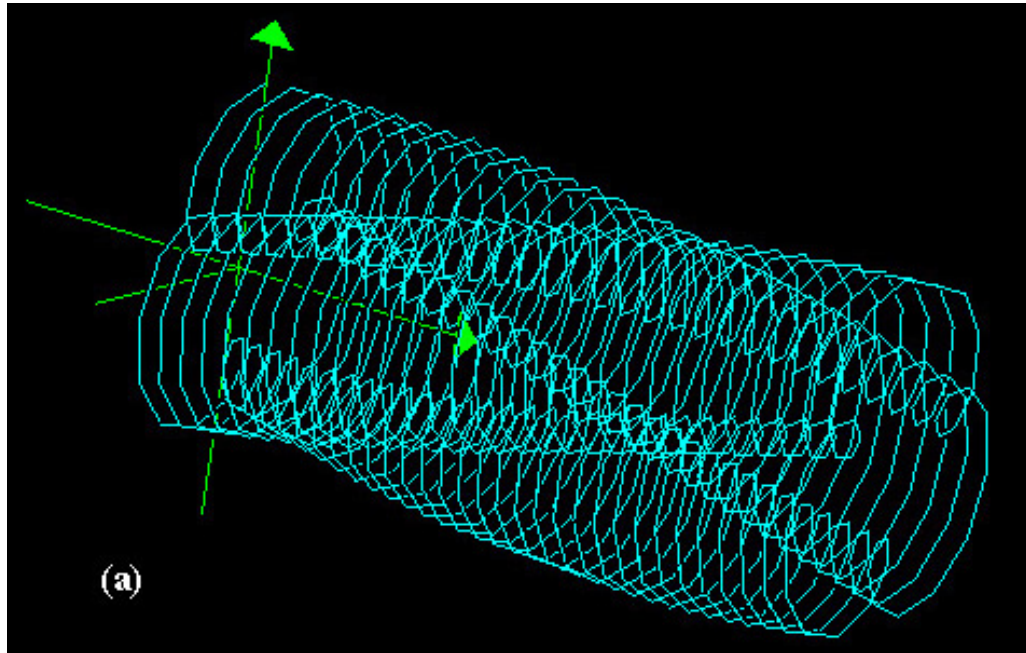
Two-dimensional signals of a mistuned Octave: $f_1 = 30\text{Hz}$; $f_2 = 61\text{Hz}$; with $A_1 = 1$ and (a) $A_2 = 1$ or (b) $A_2 = 0.5$. Although in the case of mistuned Unisons the beats are strongest for interfering (co)sines with the same amplitudes, as the ratio f_2 / f_1 increases [i.e. mistuned Octave: $f_2 / f_1 \approx 2$; mistuned Thirteenth (Octave and a Fifth): $f_2 / f_1 \approx 3$; mistuned double Octave: $f_2 / f_1 \approx 4$; etc.], strongest beats are obtained for an increasingly lower amplitude of the high-frequency tone (Plomp, 1966). Based on the proposed three-dimensional signals (see the following pages), the strongest beats are obtained when the amplitude ratio of the interfering (co)sines is inversely proportional to their frequency ratio. I.e. if $f_2 / f_1 \approx 3$ and $A_1 = 1$, then $A_2 \approx 0.3$.



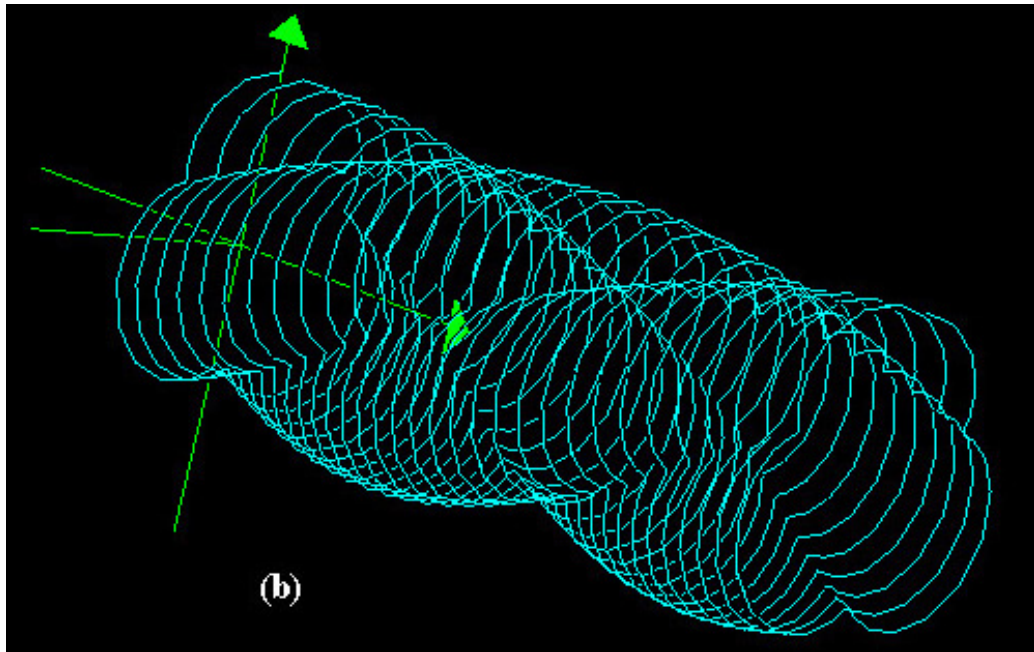
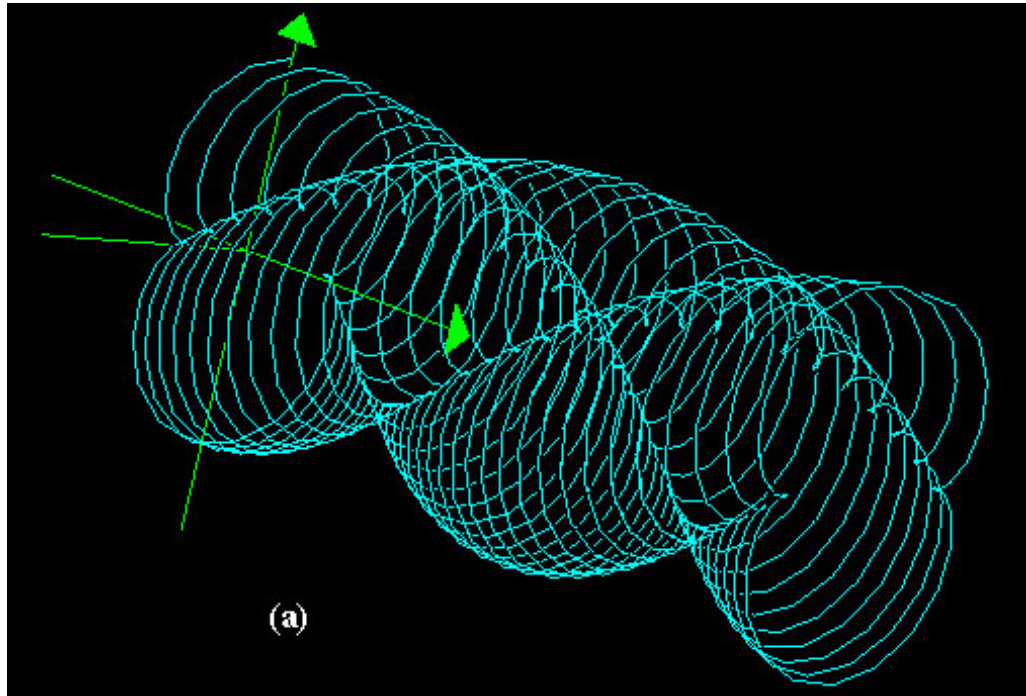
Three-dimensional signals of a mistuned Octave: $f_1 = 30\text{Hz}$; $f_2 = 61\text{Hz}$; with $A_1 = 1$ and (a) $A_2 = 1$ or (b) $A_2 = 0.5$. The amplitude fluctuations in (b) correspond to the strongest beats. In this case, the amplitude relationship of the interfering cosines results in signal twists that avoid the conflicting amplitude fluctuations present in (a).



Three-dimensional signals of a mistuned Thirteenth (an Octave and a Fifth): $f_1 = 30\text{Hz}$; $f_2 = 91\text{Hz}$; with $A_1 = 1$ and (a) $A_2 = 0.5$ or (b) $A_2 = 0.3$. Consistent with earlier studies, the amplitude of the higher frequency component, for which the signal-twists are free from conflicting amplitude fluctuations and the beats are strongest, is lower than it was in the case of the mistuned Octave [signal (b); $A_2 = 0.3 < 0.5$].



Three-dimensional signals of a mistuned Double Octave: $f_1 = 30\text{Hz}$; $f_2 = 121\text{Hz}$; with $A_1 = 1$ and (a) $A_2 = 0.5$ or (b) $A_2 = 0.2$. Again, the amplitude of the higher frequency component, for which the signal-twists are free from conflicting amplitude fluctuations and the beats are strongest, is lower than it was in the case of the mistuned Thirteenth [signal (b); $A_2 = 0.2 < 0.3$].



(a) Mistuned Thirteenth: $f_1 = 30\text{Hz}$; $f_2 = 92\text{Hz}$; $A_1 = 1$; $A_2 = 0.3$. (b) Mistuned Double Octave: $f_1 = 30\text{Hz}$; $f_2 = 122\text{Hz}$; $A_1 = 1$; $A_2 = 0.2$. In both cases, the twisted spiral signals clearly capture the amplitude fluctuation parameters corresponding to the beating of the mistuned consonances, illustrating a likely origin.

Glossary

Amplitude

The amplitude of a vibration is the maximum displacement (velocity, pressure, etc.) from a point of rest. It is represented on the y axis of a signal or spectrum and is an expression of the energy in a vibration. Amplitude is defined in terms of relative rather than absolute reference points. The reference points are selected based on the source-receiver system of interest and do not have to be fixed but can belong to another varying phenomenon. Based on such a relative definition of amplitude, pressure amplitude fluctuations of sound waves in air are characterized by a pressure amplitude value and therefore carry energy.

Amplitude Fluctuation

Amplitude fluctuation is a special case of wave modulation and the most common time-variant characteristic of musical sounds. It is a manifestation of wave interference and consists of variations in the maximum value (i.e. amplitude) of sound signals relative to a reference point. Amplitude fluctuation is a physical phenomenon described in terms of rate, degree, and shape of fluctuation and manifested perceptually as beating, roughness, or

combination tones (depending on the fluctuation rate). The terms amplitude fluctuation and amplitude modulation are used interchangeably in the literature and the assumption of their equivalence introduces numerous problems discussed in the present study.

There are four main arguments for the wave nature of amplitude fluctuations and their energy content:

- Amplitude fluctuations can transfer energy to systems (i.e. machinery, resonators, etc.) whose natural frequency matches the fluctuation rate.
- Amplitude fluctuations of acoustic waves correspond to changes in the vibration velocity of a mass at a rate equal to the rate of fluctuation, indicating consumption / radiation of energy at the same rate.
- Amplitude fluctuations entering a dispersive medium separate from the wave that carries them and travel independently at an observable rate and strength.
- Successive portions of a diffracted wave-front are determined by the envelope of secondary wavelets, indicating that amplitude fluctuations and the wave envelopes that describe them behave as waves.

Amplitude Fluctuation Rate

Amplitude fluctuation rate is the number of times the amplitude of a complex signal varies per second. For two-component complex signals, the amplitude fluctuation rate is equal to the frequency difference between the two

components. For complex signals/spectra with more than two components the fluctuation rate depends on the components' relative frequency and amplitude values and may be non-periodic.

Amplitude Fluctuation Degree (AF-degree)

Amplitude fluctuation degree is the difference between the maximum and the minimum amplitude values of a complex signal, relative to its maximum value:

$AF_{degree} = (A_{max} - A_{min}) / A_{max}$. For two-component complex signals, AF-degree depends on the relative amplitudes of the two components:

$$AF_{degree} = 2A_2 / (A_1 + A_2) \quad \text{where} \quad A_1 \geq A_2.$$

For complex signals with more than two components, AF-degree depends on the components' relative amplitude and phase values. AF-degree is qualitatively and quantitatively different from a signal's amplitude modulation

depth (AM-depth): $AF_{degree} = \frac{2AM_{depth}}{1 + AM_{depth}}.$

Whenever AM-depth values are used as a measure of a modulated signal's AF-degree, modulation implementation produces an error arising from the nonlinear relationship between the presumed and applied AF-degrees.

Amplitude Fluctuation Shape

Amplitude fluctuation shape is the shape of a complex signal's envelope and depends on the relative frequency, amplitude, and phase values of the complex

signal's components. Amplitude fluctuations of sinusoidally amplitude modulated signals have sinusoidal shape.

Amplitude Modulation

Amplitude modulation is a spectral modification process (i.e. mixing the carrier signal with the modulating signal in a nonlinear device) that produces discrete upper and lower sidebands (i.e. for sinusoidal carrier and modulating signals, the sum and difference frequencies of the carrier and the modulating signal). Sinusoidal amplitude modulation (AM) of a sine is therefore a process that introduces two additional components (sidebands) at frequencies determined by the modulation frequency f_{mod} and amplitudes determined by the modulation depth m :

$$(A + Am \cos 2\pi f_{\text{mod}} t) \sin(2\pi f t + \phi) = A \frac{m}{2} \sin[2\pi(f - f_{\text{mod}})t + \phi_1] + A \sin(2\pi f t + \phi) + A \frac{m}{2} \sin[2\pi(f + f_{\text{mod}})t + \phi_2]$$

where f_{mod} and m ($0 \leq m \leq \infty$) are the modulation parameters and

f , A , and ϕ are the parameters of the sine with $|\phi_1 - \phi| = |\phi_2 - \phi| = \pi$ rads ($\pi = 3.14$; t : time).

The envelope of the modulated signal is an analog of the modulating signal.

Amplitude Modulation Depth (AM-depth)

Amplitude modulation depth is a measure of the spectral energy spread of an amplitude modulated signal. More specifically, $AM\text{-depth} = m$ indicates that the sum of the sidebands' amplitudes will be equal to m multiplied by the amplitude of the carrier. A side-effect of the amplitude modulation process is that the amplitude of the modulated signal will fluctuate between a minimum and a maximum value. In the simple case of sinusoidal amplitude modulation of a sine signal, AM-depth can be expressed in terms of these values:

$$AM_{depth} = (A_{max} - A_{min}) / (A_{max} + A_{min}).$$

AM-depth is qualitatively and quantitatively different from the degree of amplitude fluctuation (AF-degree) of a signal and is, therefore, not a good

measure of the modulated signal's AF-degree. : $AM_{depth} = \frac{AF_{deg\ ree}}{2 - AF_{deg\ ree}}.$

(Adjusted) Amplitude Modulation Depth Coefficient

The term refers to the amplitude modulation depth coefficient m that needs to be inserted in the amplitude modulation implementation equation

[$A(t) = A + Am(\cos 2\pi f_{mod} t)$] for an intended AF-degree to be applied:

$$m = \frac{AF_{deg\ ree}}{2 - AF_{deg\ ree}}.$$

Beats / Beating

Beating is the most familiar perceptual manifestation of amplitude fluctuation. It describes loudness fluctuations perceived when sound signals with amplitude fluctuation rate $\approx \leq 20$ per second reach the ear. The number of loudness fluctuations / beats is equal to the amplitude fluctuation rate. For the simple case of beats that result from adding together two sine signals (f_1, f_2 ; $|f_1 - f_2| \approx \leq 20$), the beating rate is equal to the frequency difference $|f_1 - f_2|$ between the two signals. Similarly to roughness, the beating sensation associated with certain complex tones is understood in terms of sine-component interaction within the same critical band. Beating is linked to the Western musical concepts of consonance and dissonance, with its presence believed to be one of the acoustical cues behind dissonance. Although the two terms have been used interchangeably in the music cognition and acoustics literature, beats and amplitude fluctuations are not equivalent. The term ‘beats’ describes only one of the perceptual attributes of amplitude fluctuation.

(Dichotic) Beats

Dichotic beats are beating-like sensations reported in dichotic experiments.

Beats of Mistuned Consonances

Beats of mistuned consonances are faint loudness fluctuations accompanying slightly mistuned consonances (i.e. mistuned Octaves, Fifths, etc.). The

traditional explanation for the beats of mistuned consonances states: ‘For justly tuned consonances between complex tones, the frequencies of certain mid and high components coincide. As the consonances are detuned, the frequencies of these components start to shift, resulting in slow amplitude fluctuations that give rise to the beating sensation’. The main problem with the above explanation is that beating sensations have also been observed in the case of mistuned consonances between sines. Early explanations of this phenomenon were based on the nonlinear creation of combination tones or harmonics inside the ear. However, in his arguments against using the best beats method to detect aurally created harmonics of a sine tone, von Békésy (1960s) demonstrated that mistuned consonances of sines do give rise to beating-like sensations even when applied to the skin of the arm. This observation challenged the early explanations, leaving open the question of beating in the case of mistuned consonances between sines. From within the context of the present study, the signals of mistuned consonances between sines exhibit amplitude fluctuations at the observed beating rates and degrees and can therefore explain the beats of mistuned consonances without the need for a nonlinear generating mechanism, whether at a cochlear or a neural level.

Best Beats Method

The best beats method is an experimental method used to examine the existence of aurally created harmonics or combination tones. Subjects are

presented with a sine signal, whose frequency is slightly removed from that of the hypothetical aural harmonic / combination tone. The observation of beats is considered as evidence of the presence of the specific aural harmonic / combination tone. The amplitude of the experiment-introduced sine for which the beats sound strongest (or ‘best’) is considered equivalent to the amplitude of the hypothetical aural harmonic / combination tone, since interference of two sines with equal amplitudes results in the strongest beats.

Binaural (Experiments)

In binaural experiments, the same stimulus is presented to both ears through headphones.

Binaural Diplacusis

Binaural diplacusis is a physiologically-based phenomenon that causes a pitch difference between a person's two ears for the same incoming frequency. It results in normal discrepancies of up to ~1 whole tone, while, in pathological cases, the pitch difference between the two ears may exceed an Octave.

Binaural diplacusis in normal ears seems to not only depend on, but even originate from structural irregularities of the basilar membrane, supporting the place theory of pitch.

Bone Conduction

The term refers to cochlear stimulation caused by vibration transfer via the bones of the skull rather than the air. The sound level reaching the cochlea via bone conduction is approximately 40dB–55dB lower than the level reaching the cochlea through the ear canal, and is higher for lower frequencies. Bone conduction efficiency may be increased further due to the occlusion effect.

Combination Tones

The term was introduced by Helmholtz to describe tones that can be traced not in a vibrating source but in the combination of two or more waves originating in vibrating sources. A specific combination tone, the difference tone, is one of the perceptual manifestations of amplitude fluctuation. Combination tones are the products of wave interference and have significant physical origins along with physiological, neurological, and cognitive ones. The physical origins of combination tones can be attributed to the transfer characteristics of the wave propagation medium. For sound waves propagating in air, the physical origin of combination tones is based on the following asymmetry:

effective propagation velocity: $c_{ef} > \text{velocity of sound in air: } c_0$ in

condensations and $c_{ef} < c_0$ in rarefactions.

Complex Analytic Signal

A complex (i.e. including imaginary terms) analytic (i.e. with a spectrum that has only positive frequency components) signal $z(t)$, corresponding to a real signal $y(t)$, is a three dimensional signal deriving from mixing the real signal $y(t)$ with its Hilbert transform $H[y(t)]$ multiplied by the unit of imaginary numbers, $i = \sqrt{-1}$. So: $z(t) = y(t) + iH[y(t)]$.

Complex analytic signals are similar to the three-dimensional signals proposed by the present study and are useful to the examination of wave propagation for numerous reasons. One of them is that the 90° phase difference between a wave signal and its Hilbert transform reflects the difference between potential and kinetic energy in a propagating wave. In the above equation, for example, $y(t)$ can be seen as representing scaled pressure and $H[y(t)]$ as representing scaled velocity. The envelope $A(t)$ of a signal $y(t)$ can therefore be seen as tracing the changes in the total (potential + kinetic) instantaneous energy of the signal over time and can be described by the absolute value of the complex analytic signal $z(t)$. So: $A(t) = \sqrt{y(t)^2 + H[y(t)]^2}$.

Complex Signal

Any signal that does not have a sinusoidal shape is called complex. Fourier (1800s) proved that complex signals can be analyzed mathematically into the sum of a set of sinusoidal signals, called Fourier components or partials of a

given complex signal and making up the complex signal's spectrum. Analysis of a complex signal into sinusoidal components is called Fourier analysis, while the reverse process (i.e. constructing a complex signal out of a set of sinusoids) is called Fourier synthesis. For a periodic complex signal, the lowest frequency component is called the fundamental. The rest of the components have frequencies that are integer multiples of the frequency of the fundamental (i.e. if the fundamental has frequency f , then the components have frequencies $1f, 2f, 3f, 4f, \dots$ etc.). Such a complex signal is called harmonic; all other complex signals are called inharmonic.

Complex Tone

The term refers to the sensation arising when a complex wave with fundamental (or amplitude fluctuation component) frequency and amplitude values within the auditory limits reaches the ear.

Complex Wave

A complex wave originates in a complex vibration and is represented by a complex signal.

Consonance / Dissonance

Consonance and dissonance are multidimensional concepts describing the degree of pleasantness / annoyance of a sound, or the degree to which a sound

fits to other sounds within a larger musical context. The primary acoustical cue determining consonance / dissonance is the absence / presence of roughness respectively (although the opposite relationship may also hold depending on musical tradition). Within the Western musical tradition, the presence of roughness is equivalent to acoustic or sensory dissonance.

Cosine Signal

A cosine signal is a signal with sinusoidal shape and a 90^0 phase-shift from the respective sine signal. The ‘real’ portion of a sinusoidal vibration (wave) is represented by a cosine signal and the ‘imaginary’ portion by a sine signal. The combination of the real and imaginary portions is the mathematically correct representation of a wave (the solution of the general wave equation is a complex number) and results in the three-dimensional signals proposed by the present study. Sine rather than cosine signals were used for certain introductory equations of the present study to follow convention and to avoid confusion between mathematical representation (cosine) and standard terminology (sinusoidal).

Critical Band(width)

The term critical band, introduced by Fletcher in the 1940s, referred to the frequency bandwidth of the, then loosely defined, auditory filter. Since von Békésy 's studies (1960s) the term also refers literally to the specific area on

the basilar membrane (located in the inner ear, inside the cochlea) that goes into vibration in resonance with an incoming sine wave. Its length is determined by the elastic properties of the membrane and has an average value of ~1mm. Both beating and roughness are understood in terms of sine component interaction within the same critical band. Studies by von Békésy and Plomp (1960s) examining the beating and roughness of mistuned consonances indicate that these sensations arise even when the added sines are separated by frequencies much larger than the critical bandwidth. The experimental results of both studies challenge earlier explanations of this phenomenon that were based on the nonlinear creation of combination tones or harmonics inside the ear. Although their final interpretations differ, both studies link the beating and roughness sensations of mistuned consonances directly to the amplitude-fluctuations of the resulting two-component complex signal, rather than to critical band.

Cross Hearing

The term refers to stimulus transfer across ears. The major cross hearing parameter is bone conduction.

Dichotic (Experiments)

In dichotic experiments, a different stimulus is presented to each ear through headphones.

Difference Tone

The difference tone is a tone with pitch corresponding to the frequency $|f_1 - f_2|$ (i.e. amplitude fluctuation rate), heard when two tones with fundamental frequencies f_1 and f_2 are played together. There is a lot of experimental evidence indicating that the difference tone can be traced to the nonlinear response of the inner ear (cochlea). These evidence, although often interpreted as such, do not question difference tone's objective nature. They simply show that the difference tone may be reinforced due to the nonlinear characteristics of the inner ear. Many researchers have been aware of the possibility of the difference tone existing outside the human ear and have been able to detect it experimentally. Large amplitudes of the interfering sources, listed in the psychoacoustics literature as a necessary condition for the nonlinear creation of difference tones, is only necessary to the nonlinear creation of difference tones inside the ear. Collinearity of the interfering sources has been identified as a principal condition for the physical creation of the difference tone. Several issues regarding combination tones in general and the difference tone in particular have been addressed in physical acoustics through the study of scattering of sound by sound. Westervelt's theory of mutual scattering (late 1950s) treats the difference tone produced by the interference of two collinear sound beams as a pressure wave and names it 'parametric array'. The term was introduced to highlight the similarity between the difference frequency

and the ‘endfire array’ (effect related to underwater signaling techniques) in terms of a) narrow directivity at low frequencies and b) low absorption.

Dispersion

The term describes the dependence of wave propagation velocity on wave frequency. Although the air is a non-dispersive medium (i.e. sound-wave velocity in air is independent of frequency), the dispersive case is addressed in the present study for two related reasons:

- Dispersion provides a case where modulations are isolated from the waves that carry them and can therefore be studied easier (assuming that the only characteristic that changes during dispersion is the modulations’ velocity).
- Systems with dispersion provide better cases for the mathematical analysis of the kinematic properties of waves (i.e. frequency, wavelength, phase and group velocities). Such an analysis is crucial to the study of modulations, of which amplitude fluctuations are just a special case.

Introducing dispersion theory to the study of musical sound is not unusual.

The study of dispersion usually starts by examining the propagation of energy concentrations in the phenomenon of beats, indicating that linking dispersion to modulations/amplitude fluctuations is commonplace rather than the exception. With energy concentrated in two distinct frequency regions (modulation rate and carrier frequency), the phenomenon of beats provides a good example of a situation that illustrates the difference between a wave's

group and phase velocities. At the same time, the difference between group and phase velocities makes it possible to observe the energy content of amplitude fluctuations / modulations at work. In music, sound diffraction, absorption, and reverberation are examples of phenomena that have been better understood with the aid of dispersion theory. The present study illustrates the usefulness of dispersion theory to the understanding of amplitude fluctuations as waves.

Energy Conservation Law (for Vibrations)

The energy conservation law states that, in the absence of external forces, the sum of kinetic and potential energies in a vibration remains constant. In spite of this statement, the power of a complex wave is, in general, not equal to the sum of the powers of its sine components. Linearity of superposition does not apply to energy quantities of wave motion, which are quadratic with respect to field variables such as vibration velocity. However, in the case of interference, if the response time of the receiving system (i.e. air molecules, eardrum, etc.) is long relative to the period of the complex wave's amplitude fluctuations, the power of the superposition is equal to the sum of the powers of the individual components.

(Signal) Envelope

The envelope of a two-dimensional complex signal is a boundary curve that traces the signal's amplitude through time. It encloses the area outlined by all maxima of the motion represented by the two-dimensional signal and includes points that do not belong to the signal. A signal's envelope is not, therefore, a member of the family of sine curves forming the complex two-dimensional signal and will consequently not be detected through FFT analysis. An envelope can be considered as the signal of the amplitude fluctuations of a modulated (or any complex) signal. Three-dimensional signals such as the ones introduced by the present study illustrate that amplitude fluctuations and the signal envelopes that describe them are not just boundary curves but waves that trace changes in the total instantaneous energy of a signal over time, representing the oscillation of potential and kinetic energies within a wave. Huygens's principle provides additional evidence supporting the claim that wave envelopes behave as waves.

Equal Temperament (Equally Tempered Tuning)

In equal temperament, the Octave is divided into 12 modular units (semitones) and all other intervals in a scale are constructed based on this unit. An interval of a semitone between 2 notes represents a fixed ratio ($=2^{1/12} \approx 1.059$) between 2 frequencies. In equal temperament only the Octave, the Fourth, and the Fifth intervals are expressed by simple number ratios.

Frequency Zoom Analysis

Frequency zoom analysis is a spectral analysis technique, often based on the Goertzel algorithm, estimating the frequency amplitudes of a complex signal for a frequency range of interest. The Goertzel algorithm calculates the Discrete Fourier Transform (DFT) of a data set $y(n)$ using an Infinite Impulse Response filtering operation. When all frequency lines are required, frequency zoom analysis is inefficient relative to Fast Fourier Transform (FFT) analysis. However, it is very efficient when only one or few frequency lines are required. Like all frequency analysis techniques, frequency zoom analysis introduces spectral spreading or ‘leakage’ that results from violating the periodicity and infinite-length assumptions for the data records analyzed. Moreover, the leakage is not symmetrical around each frequency component because of the leakage contributions of positive and negative frequencies into each other. Windowing of the data reduces but does not eliminate leakage. For accurate estimation of the frequency amplitudes in a complex signal, DFT analysis must be combined with leakage correction methods more sophisticated than simple averaging over several analysis bands centered on a frequency of interest.

(Tonal) Fusion

The fusion of two or more simultaneous tones is proportional to the degree to which the tones are heard as a single perceptual unit. Tonal fusion occurs when simultaneous complex tones contain coincident or complimentary components, consistent with the possible existence of a single complex tone. According to some researchers, fusion is the basis of musical consonance.

Ganga

Ganga is a style of singing common in Bosnia-Herzegovina and the Dalmatian Zagora regions of the Balkans. *Ganga* songs consist of two alternating sections, one sung by a soloist and one by a soloist and a chorus (three to five singers). The melodic range rarely exceeds a Fourth while, in the choral sections, voices sing at Minor / Major 2nd intervals that may or may not alternate with Unison passages. Singers must sing loudly and maintain uniform strength between each other and through time. The voices must be as identical as possible, nasal, and without vibrato so that they blend. These loudness and timbral requirements are accompanied by very specific performance arrangements and are similar to the conditions encouraging beating and roughness. Singers never move or dance while singing. They stand very tightly together, in an arch, turned slightly towards each other so that their voices will ‘collide’ at the right point. Within its geographical territory, *ganga* is valued for its distinct sonic effect rather than its semantic

content and provides a striking example of a musical genre that relies heavily on the sonic effects produced through the manipulation of amplitude fluctuation.

Gongs

Gongs are idiophones made of metal shells. Gongs of various sizes and shapes are the backbone of Indonesian orchestras (gamelan). The mass, stiffness and thickness of the gong-shells play an important role in the tuning of the instrument's vibrational modes. Common to gongs of all sizes are: a) a curved rim that creates an impedance mismatch between face and rim, essentially fixing the edges for most vibrational modes and b) a protruding (or sometimes sunken) dome/boss at the center, whose load is responsible for the Octave relationship between the two axially symmetric modes that are excited first and are associated with the gong's pitch. One of the main differences between gongs and other metallophones, such as bells, is that most of the vibratory energy of gongs emanates from the center rather than the edges of the instrument. Fixing the edges facilitates the maintenance and control of beating effects within a single instrument. The intentional incorporation of specific beating effects in a variety of musical contexts indicates that, for the Javanese and Balinese musical traditions, the beating sound of gongs has both sonic and cultural significance.

Hilbert Transform Demodulation (Envelope Extraction)

If $y(t) = A(t)\cos(2\pi ft)$ is a real signal with frequency f and slowly varying amplitude $A(t)$, its Hilbert transform is $H[y(t)] = A(t)\sin(2\pi ft)$. The Hilbert Transform provides a 90° phase shift to its input and can be thought of as the imaginary component of the real signal $y(t)$. When combined with the real signal, it results in a complex analytic signal of the type $z(t) = y(t) + iH[y(t)]$ ($i = \sqrt{-1}$). The amplitude envelope $A(t)$ can be extracted by taking the absolute value of the complex analytic signal: $A(t) = \sqrt{y(t)^2 + H[y(t)]^2}$. The absolute amplitude level obtained from spectral analysis of an envelope signal has no meaning. Nonetheless, the amplitude peaks in the spectrum of an envelope are often used as an indication of the envelope's power when compared to the mean noise floor. An example is the use of envelope extraction techniques to diagnose faults in mechanical gears (i.e. the relative amplitudes in the envelope spectrum of a faulty gear's signal indicate a fault's severity). For this reason, although absolute amplitude levels of the envelope spectrum are meaningless, they may provide a measure of envelope power when compared to the levels of another component, within the same analysis, with known power.

Huygens's Principle

According to Huygens's principle, successive portions of a wave-front are understood as deriving from the envelope of secondary wavelets. At any given moment, each point on a wave-front is considered a source of secondary wavelets that move away from it. At a later moment, the wave-front is the envelope of the combined secondary wavelets. Huygens's principle is used to explain the phenomenon of diffraction, it indicates that wave envelopes behave as waves, and provides additional theoretical evidence for the energy content of amplitude fluctuation.

Interference Principle

The interference principle is related to the superposition principle. It states that the combined amplitude of two or more vibrations (waves) may be larger (constructive interference) or smaller (destructive interference) than the amplitude of the individual vibrations (waves), depending on their phase relationship. When combining two or more vibrations (waves) with different frequencies, their periodically changing phase relationship results in periodic alterations between constructive and destructive interference giving rise to the phenomenon of amplitude fluctuation.

Just (Pure) Tuning

In just tuning, most common intervals (i.e. Fifth, Fourth, Major and Major 3rds) between the notes in a scale are expressed by the simplest possible number ratios.

Linear Superposition Principle

The principle of linear superposition, first introduced by Bernoulli (late 1700s), came out of the study of strings, considered as one-dimensional systems. It states that many sinusoidal vibrations can co-exist on a string independently of one another and that the total effect at any point is the algebraic sum of the individual motions. Of course, if the compound vibration at any point of a string is the algebraic sum of the individual vibrations, it is the sinusoidal waves (and not the vibrations) that can co-exist on a string independently of one-another. The way the principle of superposition was first stated reflects the problematic assumption of equivalence between vibrations and waves addressed in the present study. Objections against the idea of amplitude fluctuations carrying energy have occasionally been based on this principle, a principle intended for vibrations in one dimension and not waves in three. Linearity of superposition applies only to linear combinations of field variables (i.e. velocity) and not to energy quantities such as intensity or power, which are quadratic with respect to field variables.

Mijwiz

The *mijwiz* is an aerophone made of two identical cylindrical cane pipes, each with a single reed, bound together and played simultaneously. The instrument has a stepped shape. Each of the two pipes is made by joining together three cane tubes of increasing diameter: one that has the vibrating reed, one that acts as a junction, and one that carries the tone-holes. In its construction and performance practice, the *mijwiz* is an example of an instrument that makes explicit use of the perceptual richness of amplitude fluctuation, through creative exploitation of the beating and roughness sensations.

Mistuned Consonances

Mistuned consonances are intervals with frequency ratio slightly removed from a small integer ratio (i.e. slightly mistuned Octaves, Fifths, Fourths, etc.).

Modulation

The term was introduced in the acoustics and psychoacoustics literature from radio engineering to describe relatively slow variations of a signal's parameters (amplitude, frequency, and/or phase). More generally, modulations describe distortions of any arbitrary wave profile that become pronounced only on time intervals much longer than the period characteristic to the wave in question (called carrier wave). Modulations may be caused by the properties of the wave propagation medium or by the initial or boundary conditions. Seen from

the reference point of the carrier, modulations can be considered waves rather than wave-distortions. They propagate in the same medium as the individual components of a complex wave with the group velocity v_{gr} distinguished from the velocity of the components called phase velocity v_{ph} . When modulations propagate in dispersive media, they separate from the complex wave they belonged to and travel independently carrying energy, similarly to the rest of the frequency components of the complex wave. Although this will not happen in a non-dispersive medium such as air, the dispersive case serves to illustrate that, modulations in general and amplitude fluctuations in particular behave as waves in all respects and can therefore be considered as energy carriers.

Monaural (Experiments)

In monaural experiments, stimuli are presented to a single ear through headphones.

Occlusion Effect

The term refers to a marked increase in the efficiency of bone conduction for frequencies below 2000Hz, when the ear canals are occluded with earmuffs or headphone cups.

Phase

The phase of a vibration is its position within the vibration cycle at a given time. The phase of a wave is different for adjacent points in space because the vibration energy propagating in the form of a wave reaches these points at different times. Like amplitude, phase is defined in terms of relative rather than absolute reference points. The reference points are selected based on the source-receiver system of interest and do not have to be fixed but can belong to another varying phenomenon.

Pitch of the Missing Fundamental

The term refers to the observation that the pitch of a harmonic complex tone matches in frequency the fundamental component, even if this component is not actually present in the tone's original spectrum.

Pitch-Shift Effect

The term pitch-shift effect is used in the present study to refer collectively to two phenomena that are often considered discrete: the first and the second pitch-shift effects. The first pitch-shift effect describes changes in the pitch of a complex harmonic signal, resulting from the uniform frequency shift of all its components. The second pitch-shift effect describes a slight drop / rise in pitch when the frequency spacing between the components of a harmonic signal is

slightly increased / decreased respectively, while keeping the center frequency fixed.

Place Theory of Pitch

The place theory of pitch perception suggests that, in the inner ear, different points on the basilar membrane undergo maximum displacement as a function of frequency. The neurons at the point of maximum displacement specify a particular frequency, based on their position on the basilar membrane. For a pure tone, the excitation pattern (i.e. the spatial distribution of mechanical or neural activity) along the basilar membrane is limited to one small, specified area (the critical band) with maximum excitation at a single position corresponding to the particular frequency. For a complex tone there is a pattern of excitations corresponding to the tone's frequency components. Modifications to this theory are necessary in order to account for the phenomenon of the missing fundamental (or residue pitch) and for the pitch-shift effect.

Residue Pitch

The term was introduced by Schouten (1960s) as part of his periodicity theory of pitch. The theory is based on the periodicity properties of the non-resolved or 'residue' high-frequency components of a complex signal and was formulated in efforts to explain the pitch of the missing fundamental.

Roughness

The term was introduced in the acoustics and psychoacoustics literature by Helmholtz (1885) to label harsh, raspy, hoarse sounds. It refers to a harshness perceived when sound signals with amplitude fluctuation rate between ~20 and ~75-150 fluctuations per second reach the ear. Depending on the rate of fluctuation, three 'shades' of roughness have been distinguished (von Békésy, 1960s). Approximately 50 fluctuations per second give roughness of an intermediate character, lying between that of lower rates ("R" character) and that of higher rates ("Z" character). Similarly to beating, roughness is understood in terms of sine-component interaction within the same critical band. The roughness sensation is not necessarily associated with combinations of notes. It can also arise from the interference among the components of a single complex tone or from performance practices in monophonic music (i.e. fast vibrato). The only condition is that the resulting complex signal exhibits amplitude fluctuations within the specified range of fluctuation rates.

Roughness is one of the perceptual manifestations of the energy content of amplitude fluctuation and can be considered as one of the main auditory attributes along with pitch, loudness, and timbre. It can be manipulated by controlling the fluctuation rate and degree, providing means of sonic variation and musical expression. The sensation of roughness is linked to the Western musical concepts of consonance and dissonance and is the principle measure of sensory dissonance.

Signal

In acoustics, a signal is a two-dimensional graphic representation of a vibration (wave), plotting displacement (i.e. distance from a point of rest, an equilibrium, or an arbitrarily selected reference point) or change in a number of other variables such as velocity, pressure, etc. (y axis), over time (x axis). The assumed equivalence between vibrations and waves is evident in the fact that they both share the same signals. Two-dimensional wave signals have the following drawbacks: a) they do not represent wave energy quantities consistently across frequencies, b) they cannot account for the alternating positive and negative amplitude values of modulated waves with AM-depth $\geq 100\%$, c) they do not represent the amplitude fluctuation parameters, and d) they do not offer any insight regarding sound-wave propagation in air.

Simple Harmonic Motion

Simple harmonic motion is the simplest type of vibration, similar to a simple pendular motion. In a simple harmonic motion, the restoring force is proportional to the displacement (change) from a reference point of interest.

Sine Signal

A sine signal is a signal with a sinusoidal shape, representing the simplest type of vibration called simple harmonic motion.

Sine Wave

A sine wave is a wave originating from a simple harmonic motion and represented by a sine signal.

Sine (Simple/Pure) Tone

Sine tone is the sensation arising when a sine wave with frequency and amplitude values within the auditory limits reaches the ear. It is similar to the sound produced by a tuning fork or an electric sine-wave generator.

Spectrum

Spectrum is a two-dimensional graphic representation of a signal indicating the frequency (x axis) and amplitude (y axis) values of its sine components.

(3-D) Spiral Sine Signals and Twisted Spiral Complex Signals

The three-dimensional signal representation proposed by the present study is based on the complex equation of motion and includes both the sine (imaginary) and cosine (real) terms describing a vibration. It results in spiral sine signals and twisted-spiral complex signals, similar to complex analytic

signals. Projecting such signals onto two dimensions (on the real plane) results in the traditional two dimensional signal. A sine wave $y(t)$ with amplitude A and angular frequency ω is represented in two dimensions as:

$$y(t) = A \cos(\omega t) \text{ and in three dimensions as: } y(t) = A \cos(\omega t) + iA \sin(\omega t),$$

where $i = \sqrt{-1}$. Three-dimensional signals illustrate that amplitude fluctuations and the signal envelopes that describe them are not just boundary curves but waves that trace changes in the total instantaneous energy of a signal over time, representing the oscillation of potential and kinetic energies within a wave. Three-dimensional spiral sine signals offer a consistent measure of the energy content of sine waves across frequencies, while twisted spiral complex signals account for the negative amplitude values predicted by the mathematical expression of amplitude fluctuation, map the parameters of amplitude fluctuation onto the twisting parameters, and paint a more realistic picture of sound wave propagation in air.

Tambura

The *tambura* is an unfretted, long-necked lute with four strings (tuned to Octaves and a Fifth, a Fourth, or a Seventh, depending on tuning set), used to provide the drone accompaniment for one or more melodic instruments. A cotton thread (*juari*: life giving thread) is inserted between the bridge and the strings, giving the modulation of the string-length (as the string wraps/unwraps

around the bridge) a discontinuous character and resulting in the characteristic, for the instrument, buzzing sound. The buzzing is accentuated by the fact that high-frequency waves (associated with the abrupt changes in string slope and the ‘whipping’ of the bridge) propagate on the string faster than lower-frequency waves (dispersion due to string stiffness), reinforcing the energy shift to upper components and increasing the ‘brilliance’ of the buzz. Beating and roughness effects are prominent in the *tambura* sound, providing one more example of the musical relevance of amplitude fluctuation.

(Phase) Velocity

Phase velocity is the propagation velocity of a complex wave’s frequency components. For sound waves in air, phase velocity (v_{ph}) = group velocity (v_{ph}) = speed of sound (c).

(Group) Velocity

Group velocity is the propagation velocity of a complex wave’s modulations (i.e. amplitude fluctuations). Group and phase velocities are related by the expression: $v_{gr} = v_{ph} + \sigma \frac{dv_{ph}}{d\sigma}$ where $\sigma = 1/\lambda$ (λ : wavelength) is the wave index that describes waves in space just as frequency $f = 1/T$ (T: period) describes waves in time. Group and phase velocities are two distinct physical variables that, in the case of non-dispersive media (i.e. air, where

$dv_{ph}/d\sigma = 0$), have the same value. This does not mean that modulations or fluctuations lose their identity and carry no energy when they propagate in air. It simply means that, in air they propagate with the same velocity as the carrier. Regardless of whether a medium is non-dispersive or dispersive, modulations / beats / envelopes always travel with the group velocity as does wave energy.

Vibration

A vibration is a back-and-forth motion around a point of rest or, more generally, a variation of any physical property of a system around a reference value of interest. In general, observations on vibrations are expected to hold as they are for waves and observations on waves are expected to have a direct source on vibrations, expectations that have been questioned since Rayleigh's work (late 1800s). Waves originate in but are not equivalent to vibrations. Waves depend on the propagation medium properties while vibrations do not.

(Sound) Waves (in Air)

The term wave is often understood intuitively as the transport of disturbances in space, not associated with motion of the medium occupying this space as a whole. In a wave, the energy of a vibration is moving away from the source in the form of a disturbance within the surrounding medium. However, this

notion is problematic for a standing wave (i.e. a wave on a string), where energy is being transformed rather than moving, or for electromagnetic / light waves in a vacuum, where the concept of medium does not apply. The term wave implies three general notions: vibrations in time, disturbances in space, and moving disturbances in space-time associated with the transfer / transformation of energy. Based on these notions, the following is an origin-specific definition of sound waves in air: Sound waves in air represent a transfer of vibratory energy characterized by: i) rate (frequency), ii) starting position (phase), and iii) magnitude (amplitude) of vibration. In general, amplitude can be expressed equivalently in terms of maximum displacement, velocity, or pressure relative to a reference value. Sound waves in air are manifested as alternating air-condensations and rarefactions that spread away from the vibrating source with a velocity usually not related to the velocity amplitude of the vibration. They result in pressure/density disturbance patterns in the surrounding medium, which, in general, correspond to the signal that plots the vibration of the source over time.

Index of Authors¹²¹

- Arnold, S., 105.
 Ball, V., (95).
 Becker, D., (85).
 Becker, J., 15, 16.
 von Békésy, G., 4, 8, 59, 60, 83, 84, 86,
 88, 90, 94, 105-107, 116, 117, 145,
 207, 226, 242.
 Berger, E. H., (107).
 Bergman, A., 87.
 Bernstein, L. R., 109.
 Beurmann, A. E., (14).
 Beyer, R. T., 64, 94, 96-98, 177.
 Bigand, E., 86-88.
 Bilsen, F. A., (201).
 Blauert, J., 108, 160.
 de Boer, E., 46, 99-101.
 Bonett, D. G., (112, 152-154).
 Brecht, M-L., (112, 152-154).
 van den Brink, G., 46, 70, 100, 106, 116,
 (201).
 Burgess, J., C., 187.
 Burkard, R., (95, 105).
 Buunen, T. J. F., 201.
 Campbell, M., 47, 55, 104, 174.
 Cardozo, B. L., (46), 84, (100, 103).
 Carlton, M. V. Jr., 52, 53, 56-58, 83, 93.
 Carterette, E. C., 9, 14.
 Chertoff, M. E., (95).
 Crawford, F. S. Jr., 78, 170.
 Crichton, D. G., 97.
 Crowder, R. G., (221).
 Daniel, P., 204.
 Davies, J. B., 221.
 Davis, A., 177.
 Deutsch, D., 102.
 Dibben, N., 86, 87.
 Dolphin, W. F., 95.
 Elmore, W. C., 78, 82.
 Faulkner, A., 95.
 Federal Standard 1037C, 22, 80.
 Feldman, M., 180, 188.
 Fenney, M. P., 109.
 Festen, J. M., (201).
 Feynman, R. P., 76.
 Fineberg, J., (85).
 Fletcher, N. H., 10, 167, 169.
 Flottorp, G., (88).
 Fourcin, A., (95).
 Fujiwara, K., (30, 46).
 von Gierke, H. E., (107).
 Greated, C., (47, 55, 104, 174).
 Haag, R. (85).
 Hafter, E. R., (201, 203).
 Hall, D. E., 47.

¹²¹ Page numbers in parentheses indicate that the author cited is a co-author in a multi-authored study.

- Hamilton, M. F., 97, 165, 236.
Harris, C. M., 177.
Harshberger, M. L., 14.
Hartmann, W. M., (90, 91), 108, 109,
137, 144, 145, 160, (198).
Hashimoto, T., 85.
Hatano, S. (85).
Heald, M. A., (78, 82).
Helmholtz, H. L. F., 2-4, 39, 44, 59, 63,
64, 83, 84, 86-88, 92-96, 104, 196,
206, 207, 210-214, 218, 219, 222,
224, 227, 242, 243.
Hindemith, P. 83, 104.
Houtgast, T., 102.
Howard, D. M., 95.
Hoy, R. R., (108).
Hum, T. Y., (188).
Hunt, F. V., 1, 49, 52, 55, 56, 61.
Huron, D., 214.
Hutchinson, W., 39, 44, 85, 87, 88, 90,
196, 207, 210-214, 219, 223, 224,
227, 242.
Hyperception, Inc., 112, 180, 181, 183.
Imaizumi, S. 85.
Iwamiya, S., 30, 46.
Jairazbhoy, N. A., 9, (9).
Jeans, J., 94, 95.
Jenkins, J., 12.
Kameoka, A., 39, 44, 83, 86, 88, 194,
196.
Kastner, M. P., 221.
Kendall, R. A., (14), 111.
Knopoff, L., (39, 44, 85, 87, 88, 90, 196,
207, 210-214, 219, 223, 224, 227, 242).
Kohda, T., 81.
Krishnappa, G., 80, 81, 178, 179.
Kuriyagawa, M., (39, 44, 83, 86, 88, 194,
196).
Lane, C. E., 107, 145, 146, 158.
Leman, M., 204.
Levitt, H., 21.
Leighton, R. B., (76).
Lerdahl, F., (86-88).
Levelt, W. J. M., (39, 44, 83, 86-88, 194,
196, 207).
van Lieshout, R. A. J. M., (84).
Lichte, W. H., 84.
Lim, K. P., (188).
Lin, J. Y., 90, 91, 198.
Manolakis, D. G., (183, 185).
Marcuse, S., 11, 13-15.
Mason, A. C., 108.
McAdams, S., (85, 198-202, 227).
McLean, R. F., 188.
Menzel, D. H., 80.
Microsoft Corp., 111.
Moore, B. C. J., (95).
Möser, M., 177.
Musicant, A. D., (163).
Narins, P. M., 108.
Nederveen, C. J., 11.
Nixon, C. W., 107.
Okamoro, M., (30, 46).
Olsen, P. R., (12).
Ortmann, O., 84.
Oshinsky, M. L. (108).
Ostrovsky, L. A., 22, 50, 55, 78.
Pacific Tech, 249.

- Parncutt, R., (86-88).
 Perrott, D. R., 163.
 Petrović, A., 17-20.
 Pierce, A. D., 55, 61.
 Plomp, R., 3, 4, 39, 44, 52, 59, 81, 83, 84, 86-88, 93, 105, 116, 145, 194, 196, 207, 256.
 Potapov, A. S., (22, 50, 55, 78).
 Pratt, C. C., 206.
 Pressnitzer, D., 85, 198-202, 227.
 Proakis, J. G., 183, 185.
 Quante, F., (85).
 Racy, A. J., 10-12.
 Rayleigh, J. W. S., 54, 63, 94-97, 105, 108.
 Ritsma, R. J., 46, (100, 103), 232.
 Rosen, S., (95).
 Rossing, T. D., (10, 167, 169).
 Sachs, M. B., 80.
 Samuels, P., (177).
 Sands, M., (76).
 Schneider, A., 14.
 Schouten, J. F., 46, 100, 103.
 Schroeder, M. R., 46, 100.
 Seebass, T., 15.
 Sethares, W. A., 39, 44, 86, 88, 90, 105, 116, 194-196, 218, 219, 221, 225, 226, 245.
 Shera, C. A., (116).
 Smoorenburg, G. F., 46, (85), 100, 101, 104.
 StatSoft, Inc., 112.
 Stecker, C. G., 201, 203.
 Stevens, S. S., (88).
 Stoll, G., (85).
 Terhardt, E., 37, 38, 40-44, 46, 83-86, 89, 90, 196, 198, 207, 226, 232, 242.
 Towne, D. H., 50, 51, 64, 166.
 Turino, T., 221.
 Trahiotis, C., (109).
 deVale, S. C., (14).
 Vassilakis, P., 102.
 Vaughn, K., (9).
 Vogel, M., 83, 92.
 Vos, J., 85, 86, 88.
 Weber, R., (204).
 Webster, J. C., (21).
 Winsberg, S., (85).
 Wolbers, P. A., 15.
 Woodward, A. J., 112, 152-154.
 Xue, T. B., (188).
 Zeh, E., 85.
 Zweig, G., 116.
 Zwicker, E., 3, 88, 116.
 Zwislocki, J., 107.

References

- Arnold, S. and Burkard, R. (2000). Studies of interaural attenuation to investigate the validity of a dichotic difference tone recorded from the inferior colliculus in the chinchilla. *J. Acoust. Soc. Am.* 107(3): 1541-1547.
- Becker, J. (1988). Earth, Fire, *Śakti*, and the Javanese Gamelan. *Ethnomusicology* 32(3): 385-391.
- von Békésy, G. (1960). *Experiments in Hearing*. New York: Acoustical Society of America Press (1989).
- (1963a). Hearing Theories and Complex Sounds. *J. Acoust. Soc. Am.* 35(4): 588-601.
- (1963b). Three Experiments Concerned with Pitch Perception. *J. Acoust. Soc. Am.* 35(4): 602-606.
- Bergman, A. (1990). *Auditory Scene Analysis*. Cambridge MA: MIT Press.
- Bernstein, L. R. and Trahiotis, C. (1996). Binaural beats at high frequencies: Listener's use of envelope-based interaural temporal and intensive disparities. *J. Acoust. Soc. Am.* 99(3): 1670-1679.
- Beyer, R. T. (1999). *Sounds of Our Times. Two Hundred Years of Acoustics*. New York: AIP Press; Springer.
- ed. (1984). *Nonlinear Acoustics in Fluids*. New York: Van Nostrand Reinhold Company, Inc.
- Bigand, E., Parncutt, R., and Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception & Psychophysics* 58: 125-141.
- Blauert, J. (1996). *Spatial Hearing: The Psychophysics of Human Sound Localization*. [Revised edition, trans. in English by J. S. Allen]. Cambridge MA: MIT Press.

- de Boer, E. (1956a). Pitch of inharmonic signals. *Nature*, 178(4532): 535-536.
- (1956b). *On the 'Residue' in Hearing*. Doctoral dissertation. The Netherlands: University of Amsterdam.
- van den Brink, G. (1970). Two experiments on Pitch Perception: Diplacusis of Harmonic AM Signals and Pitch of Inharmonic AM Signals. *J. Acoust. Soc. Am.* 48(6/2): 1355-1365.
- Burgess, J. C. (1998). Practical considerations in Signal Processing. In *Handbook of Acoustics*, M. J. Crocker, editor: 1063-1081. New York: John Wiley & Sons, Inc.
- Buunen, T. J. F., Festen, J. M., Bilsen, F. A., and van den Brink, G. (1974). Phase effects in a three-component signal. *J. Acoust. Soc. Am.* 55(2): 297-303.
- Campbell, M. and Greated, C. (1987). *The Musician's Guide to Acoustics*. New York: Schirmer Books.
- Cardozo, B. L. and van Lieshout, R. A. J. M. (1981). Estimates of annoyance of sounds of different character. *Applied Acoustics* 14: 323-329.
- Carlton Maley, V. Jr. (1990). *The Theory of Beats and Combination Tones, 1700-1863*. [Harvard Dissertations in the History of Science. O. Gingerich, editor]. New York: Garland Publishing, Inc.
- Carterette, E. C. and Kendall, R. A. (1994). On the Tuning and stretched octave of Javanese Gamelans. *Leonardo Music Journal* 4: 59-68.
- Carterette, E. C., Kendall, R. A., and DeVale, S. C. (1993). Comparative acoustical and psychoacoustical analyses of gamelan instrument tones. *J. Acoust. Soc. Jap. (E)* 14(6): 383-396.
- Carterette, E. C., Vaughn, K., and Jairazbhoy, N. A. (1989). Perceptual, acoustical, and musical aspects of the Tambura drone. *Music Perception*, 7(2), 75-108.
- Courrech, J. (1998). Condition monitoring of machinery. In *Handbook of Acoustical Measurements and Noise Control*, C. M. Harris, editor: 39.1-39.25. New York: Acoustical Society of America Press.
- Crawford, F. S. Jr. (1973). *Berkeley Physics Course: Waves*. [Trans. in Greek by G. Voudouoglou, 1979]. Berkeley: McGraw-Hill, Inc.

- Crighton, D. G. (1998). Propagation of finite-amplitude waves in fluids. In *Handbook of Acoustics*, M. J. Crocker, editor: 187-202. New York: John Wiley & Son, Inc.
- Daniel, P. and Weber, R. (1997). Psychoacoustical roughness: Implementation of an optimized model. *Acustica* 83(1): 113-123.
- Davies, J. B. (1978). *The Psychology of Music*. Stanford: Stanford University Press.
- Davis, A. and Samuels, P. (1996). *An Introduction to Computational Geometry for Curves and Surfaces*. Oxford: Clarendon Press.
- Deutsch, D. (1999a). Grouping mechanisms in music. In *The Psychology of Music*, D. Deutsch, editor: 299-348. San Diego: Academic Press.
- (1999b). The processing of pitch combinations. In *The Psychology of Music*, D. Deutsch, editor: 349-411. San Diego: Academic Press.
- Dibben, N. (1999). The perception of structural stability in atonal music: The influence of salience, stability, horizontal motion, pitch commonality, and dissonance. *Music Perception* 16(3): 265-294.
- Dolphin, W. F., Chertoff, M. E., and Burkard, R. (1994). Comparison of the envelope following response in the Mongolian gerbil using two-tone and sinusoidally amplitude-modulated tones. *J. Acoust. Soc. Am.* 96(4): 2225-2234.
- Elmore W. C. and Heald, M. A. (1969). *Physics of Waves*. New York: Dover Publications, Inc.
- Faulkner, A., Ball, V., Rosen, S., Moore, B. C. J., and Fourcin, A. (1992). Speech pattern hearing aids for the profoundly hearing impaired: Speech perception and auditory abilities. *J. Acoust. Soc. Am.* 91(4): 2136-2155.
- Federal Standard 1037C (1996). *Telecommunications: Glossary of Telecommunication Terms*. National Communications System: Technology and Standards Division. Arlington, VA; General Services Administration: Information Technology Service. [<http://www.its.bldrdoc.gov/fs-1037>].
- Feeney, M. P. (1997). Dichotic beats of mistuned consonances. *J. Acoust. Soc. Am.* 102(4): 2333-2342.
- Feldman, M. (1994). Nonlinear system vibration analysis using Hilbert Transform: I - Free vibration analysis method. *Mechanical Systems and Signal Processing* 8(2): 119-127.

- Feynman, R. P., Leighton, R. B., and Sands, M. (1964). *The Feynman Lectures on Physics*. Menlo Park, California: Addison & Wesley Publishing Co.
- Fletcher, N. H. and Rossing, T. D. (1998). *The Physics of Musical Instruments* (2nd edition). New York: Springer-Verlag.
- Hall, D. E. (1980). *Musical Acoustics: An Introduction*. Belmont, California: Wadsworth Publishing Company.
- Hamilton, M. F. (1998). Nonlinear effects in sound beams. In *Handbook of Acoustics*, M. J. Crocker, editor: 221-228. New York: John Wiley & Sons, Inc.
- Harris, C. M. (1998). Vibration measurement techniques. In *Handbook of Acoustical Measurements and Noise Control*, C. M. Harris, editor: 10.1-10.20. New York: Acoustical Society of America Press.
- Harshberger, M. L. (1992). *Acoustics and Comparative Psychoacoustics of Indonesian Gong Tones*. Unpublished Dissertation. Los Angeles: University of California, Los Angeles; Systematic Musicology.
- Hartmann, W. M. (1999). How we localize sound. *Physics Today* 52(11): 24-29.
- Hashimoto, T. and Hatano, S. (1994). Roughness level as a measure for estimating unpleasantness: Modification of roughness level by modulation frequencies. *Proceedings of the Inter-Noise 94 Conference*, 887-892. Yokohama, Japan.
- Helmholtz, H. L. F. (1885). *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (2nd edition). Trans. A. J. Ellis. New York: Dover Publications, Inc. (1954).
- Hindemith, P. (1945). *The Craft of Musical Composition; Book I* (4th edition). New York: Associated Music Publishers, Inc.
- Houtgast, T. (1976). Subharmonic pitches of a pure tone at low S/N ratio. *J. Acoust. Soc. Am.* 60(2): 405-409.
- Howard, D. M. (1989). Peak-picking fundamental period estimation for hearing prostheses. *J. Acoust. Soc. Am.* 86(3): 902-910.
- Hunt, F. V. (1978). *Origins in Acoustics*. New York: Acoustical Society of America Press, (1992).

- Huron, D. (1991). Tonal consonance versus tonal fusion in polyphonic sonorities. *Music Perception*, 9(2): 135-154.
- Hutchinson, W. and Knopoff, L. (1978). The acoustic component of Western consonance. *Interface*, 7: 1-29.
- Hyperception, Inc. (2000). *Hypersignal Block Diagram 4.2, Standard Edition for Windows*. Dallas, Texas: Hyperception, Inc.
- Imaizumi, S. (1986). Acoustic measures of roughness in pathological voice. *Journal of Phonetics*, 14 (3/4): 457-462.
- Iwamiya, S. (1995). The effect of amplitude envelopes of each amplitude modulated wave on the timbre of compound tones consisting of three amplitude modulated waves. *J. Acoust. Soc. Jap. (E)* 16(1): 21-27.
- Iwamiya, S. and Fujiwara, K. (1985). Perceived principal pitch of FM-AM tones as a function of the phase difference between frequency modulation and amplitude modulation. *J. Acoust. Soc. Jap. (E)* 6(3): 193-202.
- Iwamiya, S. and Okamoro, M. (1996). The timbre of compound tones consisting of four amplitude modulated waves of different frequency regions. *J. Acoust. Soc. Jap. (E)* 17(2): 65-71.
- Jairazbhoy, N. A. (1995). *The Rāgs of North Indian Music. Their Structure and Evolution (2nd edition)*. Bombay, India: Popular Prakashan.
- Jeans, J. (1937). *Science and Music*. New York: Dover Publications, Inc. (1968).
- Jenkins, J. and Olsen, P. R. (1976). *Music and Musical Instruments in the World of Islam*. London: World of Islam Festival Publishing Company Ltd.
- Kameoka, A. and Kuriyagawa, M. (1969a). Consonance theory, part I: Consonance of dyads. *J. Acoust. Soc. Am.* 45(6): 1451-1459.
- (1969b). Consonance theory, part II: Consonance of complex tones and its calculation method. *J. Acoust. Soc. Am.* 45(6): 1460-1469.
- Kastner, M. P. and Crowder, R. G. (1990). Perception of the major/minor distinction: IV. Emotional connotations in young children. *Music Perception*, 8: 189-201.
- Kendall, R. A. (2001). *Music Experiment Development System (MEDS) 2001B for Windows*. Los Angeles: University of California Los Angeles, Department of Ethnomusicology, Program in Systematic Musicology.

- Kohda, T. (1985). An argument against [the] rejection of [the] role of time intervals in pitch [perception] – Revised fine structure theory of pitch perception. *J. Acoust. Soc. Jap. (E)* 6(2): 79-88.
- Krishnappa, G. (1998). Machinery condition monitoring. In *Handbook of Acoustics*, M. J. Crocker, editor, 701-711. New York: John Wiley & Sons, Inc.
- Lane, C. E. (1925). Binaural beats. *Physics Review* 26: 401-412.
- Leman, M. (2000). Visualization and calculation of the roughness of acoustical musical signals using the synchronization index model (SIM). *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*: DAFX 1-DAFX 6. Verona, Italy.
- Levitt, H. and Webster, J. C. (1998). Effects of noise and reverberation on speech. In *Handbook of Acoustical Measurements and Noise Control*, C. M. Harris, Editor: 16.1-16.20. New York: Acoustical Society of America Press.
- Lichte, W. H. (1941). Attributes of complex tones. *Journal of Experimental Psychology* 28: 455-480.
- Lin J. Y. and Hartmann, W. M. (1995). Roughness and the critical bandwidth at low frequency. *J. Acoust. Soc. Am.* 97(5): 3274-797.
- Marcuse, S. (1975). *A Survey of Musical Instruments*. New York: Harper & Row Publishers.
- (1964). *Musical Instruments: A Comprehensive Dictionary*. New York: Doubleday & Company, Inc.
- Mason, A. C., Oshinsky, M. L., and Hoy, R. R. (2001). Hyperacute directional hearing in a microscale auditory system. *Nature* 410(6829): 686-690.
- McLean, R. F., Hum, T. Y., Lim, K. P., and Xue, T. B. (1990). Machine Condition Monitoring. *Proceedings of the Institution of Engineers and Shipbuilders Conference*, Paper 1510. Scotland.
- Menzel, D. H. (1961). *Mathematical Physics*. New York: Dover Publications, Inc.
- Microsoft Corp. (1998). *Visual Basic 6.0 for 32 Bit Windows Development*. Redmond, Washington: Microsoft Corp.
- Möser, M. (1998). Vibration measurement and instrumentation. In *Handbook of Acoustics*, M. J. Crocker, editor: 689-700. New York: John Wiley & Sons, Inc.

- Narins, P. M. (2001). Acoustics: In a fly's ear. *Nature* 410(6829): 644-645.
- Nederveen, C. J. (1998). *Acoustical Aspects of Woodwind Instruments* (2nd edition). DeKalb, Illinois: Northern Illinois University Press.
- Nixon, C. W. and von Gierke, H. E. (1959). Experiments on the bone-conduction threshold in a free sound field. *J. Acoust. Soc. Am.* 31(8): 1121-1125.
- Nixon, C. W. and Berger, E. H. (1998). Hearing protection devices. In *Handbook of Acoustical Measurements and Noise Control*, C. M. Harris, editor, 21.1-21.24. New York: Acoustical Society of America Press.
- Ortmann, O. (1922). The sensorial basis of music appreciation. *Journal of Comparative Psychology*, 2: 227-256.
- Ostrovsky, L. A. and Potapov, A. S. (1999). *Modulated Waves, Theory and Applications*. Baltimore: The Johns Hopkins University Press.
- Pacific Tech (1998). *NuCalc Version 2.0 for Windows*. Woodside, California: Pacific Tech.
- Perrott, D. R. and Musicant, A. D. (1977). Rotating tones and binaural beats. *J. Acoust. Soc. Am.* 61(5): 1288-1292.
- Petrović, A. (1977). *Ganga: A Form of Traditional Rural Singing in Yugoslavia*. Unpublished Dissertation. Belfast: University of Belfast; Department of Social Anthropology.
- Pierce, A. D. (1989). *Acoustics: An Introduction to Its Physical Principles and Applications* (2nd edition). New York: Acoustical Society of America Press.
- (1998). Mathematical Theory of Wave Propagation. In *Handbook of Acoustics*, M. J. Crocker, editor: 21-38. New York: John Wiley & Sons, Inc.
- Plomp, R. (1964). The ear as a frequency analyzer. *J. Acoust. Soc. Am.* 36(9): 1628-1636.
- (1965). Detectability threshold for combination tones. *J. Acoust. Soc. Am.* 37(6): 1110-1123.
- (1966). Beats of mistuned consonances. *J. Acoust. Soc. Am.* 42(2): 462-474.
- (1976). *Aspects of Tone Sensation*. [Series in Cognition and Perception. E. C. Carterette & M. P. Friedman, editors]. New York: Academic Press.

- Plomp, R. and Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *J. Acoust. Soc. Am.* 38(4): 548-560.
- Pratt, C. C. (1921). Some Qualitative Aspects of Bitonal Complexes. *American Journal of Psychology* 32: 490-515.
- Pressnitzer D. and McAdams, S. (1999). Two phase effects in roughness perception. *J. Acoust. Soc. Am.* 105(5): 2773-2782.
- Pressnitzer, D., McAdams, S., Winsberg, S., and Fineberg, J. (2000). Perception of musical tension for nontonal orchestral timbres and its relation to psychoacoustic roughness. *Perception & Psychophysics* 62 (1): 66-80.
- Proakis, J. G. and Manolakis, D. G. (1992). *Digital Signal Processing* (2nd edition). New York: MacMillan Publishing Company.
- Racy, A. J. (1994). A dialectical perspective on musical instruments: the East Mediterranean Mijwiz. *Ethnomusicology* 38(1): 37-58.
- Rayleigh, J. W. S. (1896). *The Theory of Sound, Vol. II* (2nd edition). New York: Dover Publications, Inc. (1945).
- Ritsma, R. J. (1962). Existence region of the tonal residue, part I. *J. Acoust. Soc. Am.* 34(9):1224-1229.
- (1963). Existence region of the tonal residue, part II. *J. Acoust. Soc. Am.* 35(8): 1241-1245.
- Sachs, M. B. (2001). History of physiological acoustics: A modeling perspective. *J. Acoust. Soc. Am.* 109 (5/2): 2350.
- Schneider, A. and Beurmann, A. E. (1993). Notes on the acoustics and tuning of gamelan instruments. In *Performance in Java and Bali: Studies of narrative, theatre, music and dance*, B. Arps, editor: 197-218. London: School of Oriental and African Studies; University of London Press.
- Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (1962). Pitch of the Residue. *J. Acoust. Soc. Am.* 34(8/2): 1418-1424.
- Schroeder, M. R. (1966). Residue Pitch: A Remaining Paradox and a Possible Explanation. *J. Acoust. Soc. Am.* 40(1): 79-81.

- Seebass, T. (1993). Co-ordination between music and language in Balinese shadow-play with emphasis on *wayang gambuh*. In *Performance in Java and Bali: Studies of narrative theatre, music and dance*, B. Arps, editor: 162-173. London: School of Oriental and African Studies; University of London Press.
- Sethares, W. A. (1998). *Tuning, Timbre, Spectrum, Scale*. London: Springer-Verlag.
- Smootenburg, G. F. (1970). Pitch Perception of two-frequency Stimuli. *J. Acoust. Soc. Am.* 48(4/2): 924-942.
- StatSoft, Inc. (2000). *Statistica 5.5 for Windows*. Tulsa, Oklahoma: StatSoft, Inc.
- Stecker, C. G. and Hafter, E. R. (2000). An effect of temporal asymmetry on loudness. *J. Acoust. Soc. Am.* 107(6): 3358-3368.
- Terhardt, E. (1974a). On the perception of periodic sound fluctuations (roughness). *Acustica* 30(4): 201-213.
- (1974b). Pitch, consonance and harmony. *J. Acoust. Soc. Am.* 55(5): 1061-1069.
- (1984). The concept of musical consonance: A link between music and psychoacoustics. *Music Perception* 1(3): 276-295.
- Terhardt, E. and Stoll, G. (1981). Skalierung des Wohlklangs von 17 Umweltschallen und Untersuchung der beteiligten Hörparameter. [Musicality scaling (or the sensorial consonance) of 17 environmental sounds and investigation of the relevant hearing parameters]. *Acustica* 48(4): 247-253.
- Towne, D. H. (1967). *Wave Phenomena*. New York: Dover Publications, Inc.
- Turino, T. (1993). *Moving Away from Silence. Music of the Peruvian Altiplano and the Experience of Urban Migration*. Chicago: The University of Chicago Press.
- Vassilakis, P. (1998). A single model explaining the first and second pitch-shift effects as alternative manifestations of a single phenomenon. *Proceedings of the 5th International Conference of Music perception and Cognition*: 107-114. Seoul, Korea: Seoul National University.
- Vogel, M. (1993). *On the Relations of Tone*. Bonn: Verlag für systematische Musikwissenschaft, GmbH [Lehre von den Tonbeziehungen, 1975. Bonn, trans. V. J. Kesselbach].

- Vos, J. (1986). Purity ratings of tempered fifths and major thirds. *Music Perception* 3(3): 221-258.
- Vos, J. and Smoorenburg, G.F. (1985). Penalty for impulse noise, derived from annoyance ratings for impulse and road-traffic sounds. *J. Acoust. Soc. Am.* 77(1): 193-201.
- Wolbers, P. A. (1993). The *seblang* and its music: aspects of an East Javanese fertility rite. In *Performance in Java and Bali: Studies of narrative, theatre, music and dance*, B. Arps, editor: 34-46. London: School of Oriental and African Studies; University of London Press.
- Woodward, A. J., Bonett, D. G., and Brecht, M-L. (1990). *Introduction to Linear Models and Experimental Design*. New York: Harcourt Brace Jovanovich, Inc.
- Zeh, E., Quante, F., Becker, D., and Haag, R. (1987). Vom Phonokardiogramm zur Phonoanalyse. *Zeitschrift für Kardiologie* 76: 357-363.
- Zweig, G. and Shera, C. A. (1995). The origin of periodicity in the spectrum of evoked otoacoustic emissions. *J. Acoust. Soc. Am.* 98(4): 2018-2047.
- Zwicker, E. (1961). Subdivision of the audible frequency into critical bands. *JASA, letters to the editor* 33(2): 248-249.
- (1990). On the frequency separation of simultaneously evoked otoacoustic emissions' consecutive extrema and its relation to cochlear traveling waves. *J. Acoust. Soc. Am.* 88(3): 1639-1641.
- Zwicker, E., Flottorp, G., and Stevens, S. S. (1957). Critical band-width in loudness summation. *J. Acoust. Soc. Am.* 29(5): 548-557.
- Zwislocki, J. (1953). Acoustic attenuation between the ears. *J. Acoust. Soc. Am.* 25(4): 752-759.