

NOTE TO USERS

This reproduction is the best copy available.

UMI[®]

RECOGNITION OF SINE WAVE MODELED CONSONANTS BY NORMAL
HEARING AND HEARING-IMPAIRED INDIVIDUALS

By

RUPA BALACHANDRAN

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

2004

UMI Number: 3144077

Copyright 2004 by
Balachandran, Rupa

All rights reserved.

INFORMATION TO USERS

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

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

UMI[®]

UMI Microform 3144077

Copyright 2004 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

©2004

RUPA BALACHANDRAN

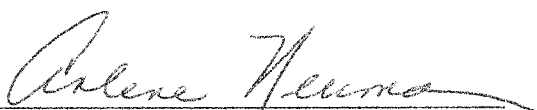
All Rights Reserved

Approval Page

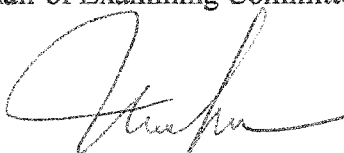
This manuscript has been read and accepted for the Graduate Faculty in Speech and Hearing Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

June 28, 2004

Date


Chair of Examining CommitteeJune 29, 2004

Date


Executive OfficerHarry LevittStuart RosenWinifred StrangePatrick Nye

Supervisory committee

Glenis Long

External Examiner

THE CITY UNIVERSITY OF NEW YORK

Abstract

Recognition of sine wave modeled consonants by normal hearing
and hearing-impaired listeners

by

Rupa Balachandran

Advisor: Professor Arlene Neuman

Sine wave modeling is a parametric tool for representing the speech signal with a limited number of sine waves. It involves replacing the peaks of the speech spectrum with sine waves and discarding the rest of the lower amplitude components during synthesis. It has the potential to be used as a speech enhancement technique for hearing-impaired adults. The present study answers the following basic questions: 1) Are sine wave synthesized speech tokens more intelligible than natural speech tokens? 2) What is the effect of varying the number of sine waves on consonant recognition in quiet? 3) What is the effect of varying the number of sine waves on consonant recognition in noise? 4) How does sine wave modeling affect the transmission of speech feature in quiet and in noise? 5) Are there differences in recognition performance between normal hearing and hearing-impaired listeners?

VCV syllables representing 20 consonants (/p/, /t/, /k/, /b/, /d/, /g/, /f/, /θ/, /s/, /ʃ/, /v/, /z/, /tʃ/, /dʒ/, /y/, /w/, /r/, /l/, /m/, /n/) in three vowel contexts (/a/, /i/, /u/) were modeled with 4, 8, 12, and 16 sine waves. A consonant recognition task was performed in quiet, and in background noise (+10 dB and 0 dB SNR). Twenty hearing-impaired listeners and six normal hearing listeners were tested under headphones at their most comfortable listening level. The main findings were: 1) Recognition of unprocessed

speech was better than that of sine wave modeled speech. 2) Asymptotic performance was reached with 8 sine waves in quiet for both normal hearing and hearing-impaired listeners. 3) Consonant recognition performance in noise improved with increasing number of sine waves. 4) As the number of sine waves was decreased, place information was lost first, followed by manner, and finally voicing. 5) Hearing-impaired listeners made more errors than normal hearing listeners, but there were no differences in the error patterns made by both groups.

Acknowledgements

I would like to acknowledge the tremendous contribution of my chair Arlene Neuman who has mentored me through all the years I have been in the doctoral program. She is an excellent role model, a wonderful teacher and a patient advisor.

Stuart Rosen is a shining example of a mentor and advisor who goes beyond the call of duty. He is a generous professional and shares his knowledge, wisdom and good humor with all around him. I consider myself extremely fortunate to have had the opportunity to work with him.

I want to thank Harry Levitt who has opened my mind to new areas and whose enthusiasm and foresight developed my interest in my dissertation topic. To be a student of Harry is to discover the joy of learning!

Pat Nye has been generous with his time and advice and his very dry wit which has kept me in good cheer. I am grateful for all the time he spent developing multiple versions of the program. I think we stopped shy of the modification that would order breakfast as well!

Winifred Strange has been the equivalent of a custom designed speech science search engine. Her mastery of her field, her ability to quickly distill problems into workable outcomes, and her tremendous enthusiasm to learn about new areas has been a source of wonder and inspiration to me.

This dissertation is testament to all that is good in the spirit of the scientific community. I am most obliged to Valerie Hazan and the wonderful team at University College London for providing me with lab space, office space, research equipment, tea bags, and guides, in order to make this study a success. I am much obliged to Andy Faulkner for spending hours making recordings and serving as a sounding board throughout the process. I am also thankful to the Royal Ear, Nose and Throat Hospital and Chelsea and Westminster Hospital for providing me with subjects for this study.

My colleagues at CSUS whose support and encouragement kept me going, my chair Jim McCartney who has provided mentoring and friendship during the difficult transition time, and my students Katrina, Anna and Sean who insisted that they enjoyed helping with thousands of spectrograms, all deserve a big thank you!

My parents and my family have helped me stay focused through what has been a very long process. Their belief in me was at many times far greater than my own, and for that I am grateful. For my nieces and nephews who kept it real.

My dear husband Vinay, whose love and support has really enabled me to stay the course and, whose golf I think will suffer greatly at the end of this dissertation!

To all of you my sincerest gratitude.

TABLE OF CONTENTS

	Page
List of Tables	xi
List of Figures	xii
List of Figures: Appendix 1	xv
List of Tables: Appendix 2	xxi
List of Tables: Appendix 3	xxiii
Chapter 1	
Introduction	1
Chapter 2	
Review of the literature	4
2.1. Acoustic and phonetic cues for speech perception	4
2.1.1. Vowels	4
2.1.2. Consonants	5
2.1.3. Temporal cues to the perception of consonants	8
2.1.4. The role of spectral information in consonant recognition	10
2.1.5. Consonant perception by hearing-impaired individuals	12
2.1.6. Investigations of the roles of audibility and effective audibility	15
2.1.7. Perception of place of articulation: Salience of short duration spectral cues	18
2.1.8. Acoustic cues to perception of voicing	19
2.1.9. Perception of fricatives	22
2.2. Performance by hearing-impaired listeners in noise	24

	Page
2.3. Signal processing to improve speech recognition of hearing-impaired listeners	27
2.3.1. Spectral sharpening	27
2.3.2. The use of sine waves to code speech	30
2.4. Evidence of improved speech intelligibility with reduced spectral information: Application with cochlear implants	31
2.5. Sine wave modeling	34
2.6. Report of early studies with the overlap-add technique	40
2.6.1. Consonant recognition	40
2.6.2. Perceptual similarity	41
2.7. Rationale	43
 Chapter 3	
Methods	45
3.1. Subjects	45
3.2. Stimuli	48
3.2.1 Sine wave modeling	48
3.3. Instrumentation	53
3.4. Procedure	55
 Chapter 4	
Results	58
4.1. Experimental results for the normal hearing group	59
4.2. Experimental results for the hearing-impaired Group	66
4.3. Observations from the confusion matrices	74
4.3.1. Quiet	75
4.3.2. +10 dB Signal-to-noise ratio	75
4.3.3. 0 dB Signal-to-noise ratio	77

	Page
4.4. Information transfer analysis	77
4.4.1. Information transfer for normal hearing adults	79
4.4.2. Information transfer for hearing-impaired adults	81
4.5. Results of the acoustic analyses	86
 Chapter 5	
Discussion	99
5.1. Number of sine waves	100
5.1.1. Acoustical analysis from spectrograms	101
5.1.2. Temporal cues	104
5.1.3. Comparison to low-pass filtering and masking	104
5.1.4. Effect of signal processing	105
5.2. Effect of signal-to noise ratio	106
5.2.1. Analysis of transmission of features	108
5.2.2. Acoustic analysis	109
5.2.3. Comparison to masking studies	109
5.3. Effect of vowel environment	111
5.4. Comparison to literature	112
5.4.1. Sine wave modeling	112
5.4.2. Comparison to cochlear implant processors	113
5.5. Comparison to spectral sharpening studies	115
5.6 Future directions for research	114
 Chapter 6	
Summary and conclusions	116

	Page
Appendix 1	
Spectrograms of selected syllables	119
Appendix 2	
Confusion matrices for normal hearing adults	190
Appendix 3	
Confusion matrices for the hearing-impaired subjects	205
References	220

LIST OF TABLES

Table	Page
1. Subject characteristics and audiometric data in the test ear for the hearing-impaired adults	46
2. Subject characteristics and audiometric data in the test ear for the normal-hearing adults	47
3. Summary table of the repeated measures analysis of variance for the normal hearing group	60
4. Mean consonant recognition performance as a function of number of sine waves, signal-to-noise ratio, and vowel context	61
5. Summary table of the repeated measures analysis of variance for the hearing-impaired group	67
6. Mean consonant recognition performance for hearing-impaired adults as a function of number of sine waves, signal-to-noise ratio, and vowel context	68
7. Abbreviations for feature matrix in the analyses	78
8. Matrix of consonants and abbreviations used to denote voicing, manner and place	78

LIST OF FIGURES

Figure	Page
1. Unprocessed /i/.	37
2. /i/ processed with 8 sine waves	37
3. /i/ processed with 4 sine waves	37
4. Short-term spectrum of cafeteria noise	50
5. Short-term spectrum of syllable /aba/	51
6. Short-term spectrum for syllable /aba/ with 4 sine waves	51
7. Short-term spectrum of /aba/ with 12 sine waves	52
8. Short-term spectrum of /aba/ with 16 sine waves	52
9. Frequency response of right earphone of Beyerdynamic DT-48 headphones	54
10. Frequency response of left earphone of Beyerdynamic DT-48 Headphones	55
11. Mean performance of normal hearing adults as a function of sine wave condition	62
12. Mean consonant recognition performance of normal hearing listeners as a function of signal-to-noise ratio	63
13. Mean consonant recognition performance of normal hearing listeners as a function of vowel context	64
14. Mean consonant recognition performance of the normal hearing group as a function of the number of sine waves and the signal-to-noise ratio	65
15. Mean consonant recognition performance of normal hearing listeners as a function of signal-to-noise ratio, vowel, and sine waves	66
16. Mean consonant recognition performance of hearing-impaired adults as a function of the number of sine waves	69

Figure		Page
17.	Mean performance of hearing-impaired listeners as a function of signal-to-noise ratio	70
18.	Mean consonant recognition performance of hearing-impaired listeners as a function of vowel context	71
19.	Consonant recognition scores of the hearing-impaired group as a function of the number of sine waves and the signal-to-noise ratio	72
20.	The mean consonant recognition performance for hearing-impaired adults in quiet as a function of the number of signal-to-noise-ratio, vowel, and sine waves	73
21.	Voicing information transferred to normal hearing adults as a function of the number of sine waves and the signal-to-noise ratio	79
22.	Manner information transferred to normal hearing adults as a function of the number of sine waves and signal-to-noise ratio	80
23.	Place information transferred to normal hearing adults as a function of the number of sine waves and signal to noise ratio	81
24.	Voicing information transferred to hearing-impaired adults as a function of the number of sine waves and signal-to-noise ratio	82
25.	Manner information transferred to hearing-impaired subjects as function of the number of sine waves and signal-to-noise ratio	83
26.	Place information transferred to hearing-impaired subjects as a function of the number of sine waves and signal-to-noise ratio	84
27.	Wideband Spectrogram on the syllable /aba/ unprocessed in quiet	87
28.	Wideband Spectrogram on the syllable /aba/ with 16 sine waves in quiet	87

a

Figure	Page
29. Wideband Spectrogram on the syllable /aba/ with 8 sine waves in quiet	88
30. Wideband Spectrogram on the syllable /aba/ with 4 sine waves in quiet	88
31. Wideband Spectrogram on the syllable /asa/ with 16 sine waves in quiet	89
32. Wideband Spectrogram on the syllable /aba/ with 16 sine waves in + 10 dB SNR	90
33. Wideband Spectrogram on the syllable /aba/ with 16 sine waves in 0 dB SNR	90

LIST OF FIGURES

Appendix 1

Figure		Page
1.	Spectrogram of /apa/ in quiet; unprocessed stimuli	119
2.	Spectrogram of /apa/ in quiet; 16 sine waves	119
3.	Spectrogram of /apa/ in quiet; 12 sine waves	120
4.	Spectrogram of /apa/ in quiet; 8 sine waves	120
5.	Spectrogram of /apa/ in quiet; 4 sine waves	121
6.	Spectrogram of /apa/ in +10 dB SNR; unprocessed	122
7.	Spectrogram of /apa/ in +10 dB SNR; 16 sine waves	122
8.	Spectrogram of /apa/ in +10 dB SNR; 12 sine waves	123
9.	Spectrogram of /apa/ in +10 dB SNR; 8 sine waves	123
10.	Spectrogram of /apa/ in +10 dB SNR; 4 sine waves	124
11.	Spectrogram of /apa/ in 0 dB SNR; unprocessed	125
12.	Spectrogram of /apa/ in 0 dB SNR; 16 sine waves	125
13.	Spectrogram of /apa/ in 0 dB SNR; 12 sine waves	126
14.	Spectrogram of /apa/ in 0 dB SNR; 8 sine waves	126
15.	Spectrogram of /apa/ in 0 dB SNR; 4 sine waves	127
16.	Spectrogram of /aba/ in quiet; unprocessed stimuli	128
17.	Spectrogram of /aba/ in quiet; 16 sine waves	128
18.	Spectrogram of /aba/ in quiet; 12 sine waves	129
19.	Spectrogram of /aba/ in quiet; 8 sine waves	129
20.	Spectrogram of /aba/ in quiet; 4 sine waves	130

Figure	Page
21. Spectrogram of /aba/ in +10 dB SNR; unprocessed	131
22. Spectrogram of /aba/ in +10 dB SNR; 16 sine waves	131
23. Spectrogram of /aba/ in +10 dB SNR; 12 sine waves	132
24. Spectrogram of /aba/ in +10 dB SNR; 8 sine waves	132
24. Spectrogram of /aba/ in +10 dB SNR; 4 sine waves	133
26. Spectrogram of /aba/ in 0 dB SNR; unprocessed	134
27. Spectrogram of /aba/ in 0 dB SNR; 16 sine waves	134
28. Spectrogram of /aba/ in 0 dB SNR; 12 sine waves	135
29. Spectrogram of /aba/ in 0 dB SNR; 8 sine waves	135
30. Spectrogram of /aba/ in 0 dB SNR; 4 sine waves	136
31. Spectrogram of /asa/ in quiet; unprocessed	137
32. Spectrogram of /asa/ in quiet; 16 sine waves	137
33. Spectrogram of /asa/ in quiet; 12 sine waves	138
34. Spectrogram of /asa/ in quiet; 8 sine waves	138
35. Spectrogram of /asa/ in quiet; 4 sine waves	139
36. Spectrogram of /asa/ in +10 dB SNR; unprocessed	140
37. Spectrogram of /asa/ in +10 dB SNR; 16 sine waves	140
38. Spectrogram of /asa/ in +10 dB SNR; 12 sine waves	141
39. Spectrogram of /asa/ in +10 dB SNR; 8 sine waves	141
40. Spectrogram of /asa/ in +10 dB SNR; 4 sine waves	142
41. Spectrogram of /asa/ in 0 dB SNR; unprocessed	143
42. Spectrogram of /asa/ in 0 dB SNR; 16 sine waves	143

Figure	Page
43. Spectrogram of /asa/ in 0 dB SNR; 12 sine waves	144
44. Spectrogram of /asa/ in 0 dB SNR; 8 sine waves	144
45. Spectrogram of /asa/ in 0 dB SNR; 4 sine waves	145
46. Spectrogram of /aza/ in quiet; unprocessed	146
47. Spectrogram of /aza/ in quiet; 16 sine waves	146
48. Spectrogram of /aza/ in quiet; 12 sine waves	147
49. Spectrogram of /aza/ in quiet; 8 sine waves	147
50. Spectrogram of /aza/ in quiet; 4 sine waves	148
51. Spectrogram of /aza/ in +10 dB SNR; unprocessed	149
52. Spectrogram of /aza/ in +10 dB SNR; 16 sine waves	149
53. Spectrogram of /aza/ in +10 dB SNR; 12 sine waves	150
54. Spectrogram of /aza/ in +10 dB SNR; 8 sine waves	150
55. Spectrogram of /aza/ in +10 dB SNR; 4 sine waves	151
56. Spectrogram of /aza/ in 0 dB SNR; unprocessed	151
57. Spectrogram of /aza/ in 0 dB SNR; 16 sine waves	152
58. Spectrogram of /aza/ in 0 dB SNR; 12 sine waves	152
59. Spectrogram of /aza/ in 0 dB SNR; 8 sine waves	153
60. Spectrogram of /aza/ in 0 dB SNR; 4 sine waves	153
61. Spectrogram of / atʃa / in quiet; unprocessed	154
62. Spectrogram of / atʃa / in quiet; 16 sine waves	154
63. Spectrogram of / atʃa / in quiet; 12 sine waves	155

Figure	Page
64. Spectrogram of / atʃa / in quiet; 8 sine waves	155
65. Spectrogram of / atʃa / in quiet; 4 sine waves	156
66. Spectrogram of / atʃa / in +10 dB SNR; unprocessed	157
67. Spectrogram of / atʃa / in +10 dB SNR; 16 sine waves	157
68. Spectrogram of / atʃa / in +10 dB SNR; 12 sine waves	158
69. Spectrogram of / atʃa / in +10 dB SNR; 8 sine waves	158
70. Spectrogram of / atʃa / in +10 dB SNR; 4 sine waves	159
71. Spectrogram of / atʃa / in 0 dB SNR; unprocessed	160
72. Spectrogram of / atʃa / in 0 dB SNR; 16 sine waves	160
73. Spectrogram of / atʃa / in 0 dB SNR; 12 sine waves	161
74. Spectrogram of / atʃa / in 0 dB SNR; 8 sine waves	161
75. Spectrogram of / atʃa / in 0 dB SNR; 4 sine waves	162
76. Spectrogram of / ana / in quiet; unprocessed	163
77. Spectrogram of / ana / in quiet; 16 sine waves	163
78. Spectrogram of / ana / in quiet; 12 sine waves	164
79. Spectrogram of / ana / in quiet; 8 sine waves	164
80. Spectrogram of / ana / in quiet; 4 sine waves	165
81. Spectrogram of / ana / in +10 dB SNR; unprocessed	166
82. Spectrogram of / ana / in +10 dB SNR; 16 sine waves	166
83. Spectrogram of / ana / in +10 dB SNR; 12 sine waves	167

Figure	Page
84. Spectrogram of / ana / in +10 dB SNR; 8 sine waves	167
85. Spectrogram of / ana / in +10 dB SNR; 4 sine waves	168
86. Spectrogram of / ana / in 0 dB SNR; unprocessed	169
87. Spectrogram of / ana / in 0 dB SNR; 16 sine waves	169
88. Spectrogram of / ana / in 0 dB SNR; 12 sine waves	170
89. Spectrogram of / ana / in 0 dB SNR; 8 sine waves	170
90. Spectrogram of / ana / in 0 dB SNR; 4 sine waves	171
91. Spectrogram of / ara / in quiet; unprocessed	172
92. Spectrogram of / ara / in quiet; 16 sine waves	172
93. Spectrogram of / ara / in quiet; 12 sine waves	173
94. Spectrogram of / ara / in quiet; 8 sine waves	173
95. Spectrogram of / ara / in quiet; 4 sine waves	174
96. Spectrogram of / ara / in +10 dB SNR; unprocessed	175
97. Spectrogram of / ara / in +10 dB SNR; 16 sine waves	175
98. Spectrogram of / ara / in +10 dB SNR; 12 sine waves	173
99. Spectrogram of / ara / in +10 dB SNR; 8 sine waves	176
100. Spectrogram of / ara / in +10 dB SNR; 4 sine waves	177
101. Spectrogram of / ara / in 0 dB SNR; unprocessed	177
102. Spectrogram of / ara / in 0 dB SNR; 16 sine waves	178
103. Spectrogram of / ara / in 0 dB SNR; 12 sine waves	178
104. Spectrogram of / ara / in 0 dB SNR; 8 sine waves	179
105. Spectrogram of / ara / in 0 dB SNR; 4 sine waves	180

Figure	Page
106. Spectrogram of / aja / in quiet; unprocessed	181
107. Spectrogram of / aja / in quiet; 16 sine waves	181
108. Spectrogram of / aja / in quiet; 12 sine waves	182
109. Spectrogram of / aja / in quiet; 8 sine waves	182
110. Spectrogram of / aja / in quiet; 4 sine waves	183
111. Spectrogram of / aja / in +10 dB SNR; unprocessed	184
112. Spectrogram of / aja / in +10 dB SNR; 16 sine waves	184
113. Spectrogram of / aja / in +10 dB SNR; 12 sine waves	185
114. Spectrogram of / aja / in +10 dB SNR; 8 sine waves	185
115. Spectrogram of / aja / in +10 dB SNR; 4 sine waves	186
116. Spectrogram of / aja / in 0 dB SNR; unprocessed	187
117. Spectrogram of / aja / in 0 dB SNR; 16 sine waves	187
118. Spectrogram of / aja / in 0 dB SNR; 12 sine waves	188
119. Spectrogram of / aja / in 0 dB SNR; 8 sine waves	188
120. Spectrogram of / aja / in 0 dB SNR; 4 sine waves	189

LIST OF TABLES

<u>Appendix 2</u>		
Table		Page
1.	Confusion matrix for normal hearing adults for unprocessed stimuli in quiet	190
2.	Confusion matrix for normal hearing adults for 16 sine waves in quiet	191
3.	Confusion matrix for normal hearing adults for 12 sine waves in quiet	192
4.	Confusion matrix for normal hearing adults for 8 sine waves in quiet	193
5.	Confusion matrix for normal hearing adults for 4 sine waves in quiet	194
6.	Confusion matrix for normal hearing adults for unprocessed stimuli at +10 dB SNR	195
7.	Confusion matrix for normal hearing adults for 16 sine waves at +10 dB SNR	196
8.	Confusion matrix for normal hearing adults for 12 sine waves at +10 dB SNR	197
9.	Confusion matrix for normal hearing adults for 8 sine waves at +10 dB SNR	198
10.	Confusion matrix for normal hearing adults for 4 sine waves at +10 dB SNR	199
11.	Confusion matrix for normal hearing adults for unprocessed speech at 0 dB SNR	200
12.	Confusion matrix for normal hearing adults for 16 sine waves at 0 dB SNR	201
13.	Confusion matrix for normal hearing adults for 12 sine waves at 0 dB SNR	202

Table	Page
14. Confusion matrix for normal hearing adults for 8 sine waves at 0 dB SNR	203
15. Confusion matrix for normal hearing adults for 4 sine waves at 0 dB SNR	204

LIST OF TABLES

<u>Appendix 3</u>		
Table		Page
1.	Confusion matrix for hearing-impaired adults for unprocessed stimuli in quiet	205
2.	Confusion matrix for hearing-impaired adults for 16 sine waves in quiet	206
3.	Confusion matrix for hearing-impaired adults for 12 sine waves in quiet	207
4.	Confusion matrix for hearing-impaired adults for 8 sine waves in quiet	208
5.	Confusion matrix for hearing-impaired adults for 4 sine waves in quiet	209
6.	Confusion matrix for hearing-impaired adults for unprocessed stimuli at +10 dB SNR	210
7.	Confusion matrix for hearing-impaired adults for 16 sine waves at +10 dB SNR	211
8.	Confusion matrix for hearing-impaired adults for 12 sine waves at +10 dB SNR	212
9.	Confusion matrix for hearing-impaired adults for 8 sine waves at +10 dB SNR	213
10.	Confusion matrix for hearing-impaired adults for 4 sine waves at +10 dB SNR	214
11.	Confusion matrix for hearing-impaired adults for unprocessed speech at 0 dB SNR	215
12.	Confusion matrix for hearing-impaired adults for 16 sine waves at 0 dB SNR	216

Table	Page
13. Confusion matrix for hearing-impaired adults for 12 sine waves at 0 dB SNR	217
14. Confusion matrix for hearing-impaired adults for 8 sine waves at 0 dB SNR	218
15. Confusion matrix for hearing-impaired adults for 4 sine waves at 0 dB SNR	219

CHAPTER 1

Introduction

Expectations from a hearing aid are fairly consistent across large sections of hearing impaired adults irrespective of age or hearing loss. Audibility of soft sounds, improvements in speech intelligibility in quiet and in noise, and comfortable levels of audibility for very loud sounds are among the top five desirable qualities for a hearing aid (Kochkin, 2002). Recent advances in hearing aid design using wide-band compression strategies are very efficient in amplifying sounds to a level that makes soft sounds audible to the listener while preventing louder sounds from causing discomfort. The effort to improve speech intelligibility has proven more challenging.

Traditional amplification rationales use an audibility based approach to hearing aid fitting. The focus is to restore audibility at frequencies where hearing loss exists. The rationale underlying this approach is that providing audibility for frequencies where hearing loss exists can improve speech intelligibility. Restoring audibility is helpful, especially in the frequencies where the hearing loss is less than 55 dB HL (Turner and Cummings, 1999). However, the effect of sensorineural hearing loss is more than a simple attenuation of specific frequencies. Hearing impairment effectively broadens the shape of the auditory filters and thereby impairs the frequency resolution of the auditory system (Moore and Glasberg, 1986, Leek and Summers, 1993). The resulting loss of frequency resolution causes a reduction in the detail of the spectral envelope of speech, thereby reducing the distinctions between frequency regions (Glasberg and Moore, 1990, Leek et al 1987). A hearing-impaired listener therefore has more difficulty than a normal

hearing listener understanding speech, especially in situations where speech has been degraded by background noise. There is thus great interest in developing methods of signal processing that will increase the signal-to-noise ratio and make speech more intelligible.

One method proposed for improving speech intelligibility is to emphasize the peaks in the spectrum of the speech signal while increasing the depth of the valleys. This method, known as “spectral sharpening” will potentially allow the impaired ear to process only the most important speech information (Bunnell, 1990, Simpson et al., 1990, Stone and Moore, 1992). A processing approach that can achieve spectral sharpening and improve signal-to-noise ratio may prove helpful in overcoming some of the degradation of the speech signal caused by spectral smearing in the impaired cochlea.

A technique akin to spectral sharpening is sine wave modeling. In this process, speech is reconstructed with the highest magnitude peaks selected from the Fast Fourier transform (McAulay and Quatieri, 1986, 1989). Reconstructing speech with only the highest magnitude peaks, and removing the rest of the speech signal has the potential to improve the signal-to-noise ratio of the speech signal (Kates, 1994). Sine wave modeling may also improve speech intelligibility in quiet, as recent information from cochlear implant studies suggests that presenting the peaks of the speech signal gives better performance than presenting information only about the fundamental frequency and F1 and F2 (Skinner et al., 1994; Wilson, Lawson and Zerbi, 1995). Thus sine wave modeling of speech may be a viable method of coding speech for the hearing-impaired.

The purpose of this study is to investigate the effect of the number of sine waves used to model speech on consonant recognition in quiet and in noise by normal hearing and hearing-impaired listeners.

CHAPTER 2

Review of the Literature

The review of the literature that follows includes information about speech as an acoustic signal, the important cues for consonant recognition by normal-hearing and hearing-impaired listeners, and signal processing techniques to enhance speech recognition of persons with hearing loss.

The speech signal contains energy in the frequency range from 50 to 10,000 Hz. The vowels typically contribute to the higher intensity periodic components with energy concentrated in the region between 250 and 3000 Hz. The consonant sounds are typically lower in intensity, have a high frequency content (650-10,000Hz) and more aperiodicity. A listener can determine the identity of speech based on the configuration of the frequency and intensity cues in the signal.

2.1 Acoustic and phonetic cues for speech perception

In order to understand how different signal processing techniques can improve speech intelligibility for hearing-impaired adults, it is helpful to review the acoustic properties important for speech recognition by normal hearing and hearing-impaired adults.

2.1.1. Vowels

Formants are regions of energy concentration in the speech spectrum generated in the vocal tract. There are three to five clearly distinguishable formants in a speech sound (F1, F2....F5). When we speak, the shape of the vocal tract changes, and the

formant frequency values change with it. The location of and the relationship between the formant frequencies are important to the correct identification of vowels. This was demonstrated convincingly when Liberman, Delattre, Cooper, and Gerstman (1954), showed that vowel identification was systematically affected by varying the formant frequencies of synthesized vowels. Many vowels can be recognized with only the two lowest frequency formants. A single formant intermediate in frequency to F1 and F2 was adequate to identify a back vowel.

2.1.2. Consonants

When the articulators move from the position for one consonant to that of the following vowel, or from a vowel to the following consonant, sudden changes in the shape of the vocal tract occur. Corresponding changes in formant frequencies are called formant transitions. Formant transitions can be very short in duration, as in the production of plosives, or long, as in the production of semivowels. Formant transitions play a very important role in the perception of consonants.

Consonants in English can be categorized by place of articulation: bilabial, labiodental, alveolar, palatal and velar, by manner of articulation: plosives, fricatives, affricates, semivowels, liquids, and nasals, and finally by presence or absence of voicing: voiced and voiceless.

The plosives are produced with a constriction at different places in the vocal tract such as the lips, teeth or soft palate. The closure in the plosives cannot be maintained and is released with a burst. In English, the plosives are /p/, /t/, /k/ and /b/, /d/, and /g/. The first three are voiceless and the latter three are voiced. Voicing starts almost immediately

after the burst is released for a voiced stop, whereas the time lag between release of the stop and voicing is longer in voiceless stops.

The acoustic cues important to perception of stop consonants can be divided into cues differentiating manner of articulation and place of articulation. The manner cues include: 1) the silence due to the articulatory closure, 2) the presence of the release burst, 3) rapid changes in the formant frequencies (especially that of the F1 transition), and 4) the rapid rise and fall time of the formants due to change of articulatory posture from vowel to stop consonant or vice-versa.

Cues for place of articulation include: 1) frequency of the release burst, 2) the onset frequency and the direction of change of the second and third formant transitions, and 3) the slope of the F2 transition. In general, a transition of the second formant with a low locus (locus: the theoretical 'throwback' to the starting point of the formant) is perceived as labial, one with a higher locus is perceived as alveolar, and with a varied vowel-dependent locus is considered palatal.

Other important distinctions in stop-consonant identification are the presence or absence of periodic vocal fold vibration (voicing) during the period of stop closure, the presence or absence of aspiration following stop release and variation in the onset time of phonation (VOT), and F1 following the release of prevocalic stops. These cues are used to distinguish between voiced and voiceless plosives.

Fricatives are characterized by aperiodic noise that is produced through a narrow constriction in the vocal tract. Like the stop consonants, they can also be characterized as voiceless and voiced. The sibilants /s/, /ʃ/, /z/, & /ʒ/ are more intense than the other fricatives and have more energy in the high frequency region between 4-6 kHz.

The labiodental and lingua-dental fricatives /f/, /θ/, /v/ and /ð/ are less intense and their spectrum is broader in frequency.

Important acoustic cues for the recognition of place of articulation of voiceless fricatives are steady-state spectral shape differences in the frication noise. While /s/ has a peak at 4-8 kHz, /ʃ/ has a peak at 2-4 kHz. The fricatives /f/ and /θ/ have similar flat spectra covering a broad frequency range. The dynamic spectral changes that occur during the formant transition between the fricative and adjacent vowel are thought to play an important role in differentiating between /f/ and /θ/. For the identification of voiced fricatives there is an additional periodic vibration or voicing along with the frication.

The production of affricates like /tʃ/ and /dʒ/ can be described as a plosive released into the neighboring fricative. The cues to their identity are similar to those of stops and fricatives.

Semivowels like /w/, /j/, /r/, and /l/ are vowel-like consonants, as they are produced with the vocal tract relatively open as in vowels. Semivowels are identified based on formant changes over time, especially the frequency change of the second formant.

Nasals like /m/, /n/ and /ŋ/ are produced with a constriction in the vocal tract at the lips, teeth or soft palate, but with simultaneous air flow to the nasal cavity produced by lowering the soft palate. Lower intensity formants and rapid frequency transitions from the vowel preceding and following nasals distinguish the place of articulation of nasals. (Borden, Harris, and Raphael, 1994)

2.1.3. Temporal cues to the perception of consonants

Information cueing the identity of a consonant is not limited to information in the spectral domain. The temporal structure of speech also provides information to the identity of the different classes of consonants. Rosen (1992) categorized temporal cues as envelope cues, periodicity cues and fine structure, depending on the fluctuation rate.

The slow fluctuation (2-50 Hz) in the overall amplitude is defined as the *envelope*. The envelope can provide information about the number of syllables or words in a signal. The envelope also contains cues that yield information about manner of articulation, and duration.

Periodicity is the periodic vibration in the temporal/amplitude domain that is the product of vocal fold vibration. It is subdivided into periodic and aperiodic information. Periodic stimulation fluctuates between 50-500 Hz and aperiodic stimulation fluctuates over a broad range from a few Hz to 5-10 kHz. In the time domain this contrast is reflected as the difference between a regularity of stimulation versus an irregularity of the speech signal. Fricatives like /s/ and /z/ have high amplitude aperiodicity as compared to vowels which are characterized by periodic vibrations. Periodicity provides voicing information about a speech sound.

Fine structure refers to changes in wave shape within single periods of periodic sounds, or over short intervals of aperiodic sounds. These fluctuations are typically between 600Hz to 10 kHz. They encode information about the spectrum of the sound including amplitude and phase information and their perception is related but not limited

to the timbre of the sound. Fine structure can provide information about the first formant frequency as the intervals between lower amplitude fluctuations.

Information in the time /amplitude domain contributes to the redundancy of cues for speech perception. In the absence of spectral cues, the category of sound can be cued by the temporal cues and vice versa. There appears to be a hierarchy of cues, with some cues to consonant identity being more robust than others when speech information is obscured by high/low pass filtering and background noise.

2.1.4. The role of spectral information in consonant recognition.

Many researchers have investigated the contribution of spectral information to speech recognition (e.g., French and Steinberg, 1946; Kryter, 1956). These studies are also relevant to an understanding of how a hearing loss (modeled as a low pass filter) might affect speech recognition performance. The present review will focus on those studies that have investigated how masking and/or filtering of speech affects consonant recognition and the transmission of phonetic/linguistic features.

Miller and Nicely (1955) were the first to carry out a systematic study of the types of consonant confusions made by normal hearing listeners under conditions of masking and filtering. Sixteen consonants were tested in a CV syllable (vowel context /a/). The stimuli were filtered with variable band pass filters (Low pass: 200, 300, 400, 600, 1200, 2500, 5000 Hz; High pass: 5000, 4000, 3000, 2500, 2000, 1000, 200 Hz) and presented at 60dB SPL. Speech was also presented with random masking noise at -18dB, -12dB, -6dB, 0, +6dB, +12dB signal-to-noise ratios. Consonants were grouped according to linguistic features. The linguistic features and the definitions used to categorize consonants were as follows:

- Voicing: presence or absence of vocal fold vibration,
- Nasality: presence or absence of soft palate closure,
- Affrication: frication in presence or absence of aperiodic noise caused by turbulence in the vocal tract,
- Duration: longer duration of consonant (s, ʃ, z, ʒ), and
- Place of articulation (front, mid, and back).

They computed an information transmission index for each of the features, indicating how much information about a feature was transmitted. Index values ranged between 0 and 1.

The transmission index as a function of signal-to-noise ratio revealed that speech features were affected differently for the different signal-to-noise ratio conditions. Raising the noise level affected transmission of information about place of articulation the most. Affrication and duration were less affected and display similar effects. Voicing and nasality were the best preserved. While place of articulation was hard to distinguish at signal-to-noise ratios poorer than +6dB, voicing and nasality were discriminable at a signal-to-noise ratio of -12dB.

The effect of low-pass filtering was similar to that of masking. Place of articulation was affected more than voicing cues. Nasality was better preserved than duration. Errors with low pass filtering were predictable depending on the high frequency energy being removed.

High pass filtering caused deterioration in performance uniformly across all linguistic features, with the exception of duration. A possible explanation is that the phonemes /s/, /ʃ/, /z/, and /ʒ/, comprising the long 'duration' group have significant high frequency energy in their spectra and this might have contributed to their robustness in perception as greater amounts of low frequency energy were removed. Error patterns with high pass filtering were dependent on the audibility of the cues for consonant perception.

The percent correct data from the filtered speech was plotted with the high-pass and the low-pass on the same graph. The frequency at which the two functions intersect

divides the frequency scale into two equivalent parts (crossover point) in terms of the importance of the frequencies below each function for speech intelligibility. By plotting the articulation score for each linguistic feature group, crossover points for each class of sound were obtained. The crossover points for features were: 450Hz for nasality, 500 Hz for voicing, 750Hz for affrication, 1900 Hz for place of articulation, and 2200Hz for duration. When the relative amount of information transmitted overall was plotted, the crossover point was 1250Hz. The analysis revealed that, at least for CV syllables, more information about speech is carried in the low frequency portion of the speech spectrum than in the high frequency regions.

To summarize, masking noise and low pass filtering degraded speech recognition performance of normal-hearing listeners in a systematic manner. Information about place of articulation and frication were most affected by masking or removing spectral information. The manner features of nasality and voicing were more resistant to masking or low pass filtering. Duration was found to be a robust feature preserved under conditions of high pass filtering. The crossover point for equivalent audibility is different for each linguistic feature. For the overall information transmitted, the crossover point is at 1250 Hz.

2.1.5. Consonant Perception by Hearing-impaired Individuals

Sensorineural hearing loss can affect the amount of acoustic information available to the listener by reducing audibility in a frequency-specific manner. Studies of the consonant confusions of hearing-impaired adults reveal that place of articulation is one of the first cues affected by hearing loss. Manner is affected next, and voicing is the

least affected of all (Bilger and Wang, 1976). Place of articulation cues tend to be those that either involve a formant transition or a change in spectral envelope. Hearing-impaired adults have tremendous difficulty processing rapid changes in formant transitions, and rely more on static spectral cues for consonant identification (Zeng and Turner 1990, Dorman et al., 1985, van Tassel et al., 1982). However, with severe hearing loss at the frequencies important for consonant recognition, the spectral envelope may be smeared due to broadening of critical bands and speech recognition is degraded. This can be readily observed in listeners with high frequency hearing losses who have significant difficulties with fricative perception.

Different configurations of hearing loss affect the audibility of the signal differently as a function of frequency. There has always been an interest in identifying patterns of consonant confusions associated with particular configurations of hearing loss. Bilger and Wang (1976) attempted to group hearing-impaired listeners according to patterns of the confusion matrices. They tested 22 hearing-impaired adults with hearing losses ranging from mild to severe with different slopes, and 3 normal-hearing adults. Stimuli were VC and CV syllables with 16 consonants in /a/, /i/, and /u/ vowel contexts presented at 40dB above speech reception thresholds.

A Sequential Information Analysis for speech (SINFA) was carried out to determine information transmitted using a variety of phonological features. They found that subjects with similar levels of performance relied on different phonological features. Based on the SINFA analysis, individuals with similar levels of overall performance were grouped roughly into three categories depending on the amount of information transmitted for each speech feature.

Group A: perceived the features of sibilance, nasality and high/anterior (place of articulation), frication and voicing. Group B: perceived the features of sibilance, duration, voicing, high/anterior, continuance and nasality. Group C: perceived the features of nasality, voicing, and had poor performance for sibilants.

Based on the word recognition scores it was possible to predict the audiogram configuration. Group A participants had normal-hearing or mild hearing loss below 1000Hz, with normal to slightly elevated speech reception thresholds and mildly reduced word discrimination scores. Group B had individuals with moderate to severe hearing losses, with flat or rising audiograms, elevated speech reception thresholds and poor word discrimination scores (50%). Group C participants had high frequency or severe hearing losses at all frequencies. Their speech reception thresholds were between the thresholds for groups A and B. They also had the poorest word discrimination scores. This group showed a large amount of variability in their performance. Patterns of information transfer scores were able to identify members for group C only, and overall performance was not a good predictor of group membership.

Similar studies looking at the relationship between consonant confusions and audiometric configurations have found that as the slope of the hearing loss increases the overall errors in consonant identification increase. As the degree of hearing loss increases, one of the first features to be affected is place of articulation followed by sibilance. High frequency hearing loss affects sibilants the most. Nasality, voicing and duration features remain intelligible even with severe hearing losses. Analysis of error patterns showed that for the plosives the substitutions were of the same voicing group with occasional voiced fricatives. Fricatives showed substitution with plosives of the

same voicing category. The voiceless affricate (/tʃ/) was substituted by either voiceless stops (/p/, /t/, /k/) or by voiceless fricatives (/s/, /ʃ/). The voiced affricate (/dʒ/) was substituted by voiced plosives (/b/, /d/, /g/). Voicing seems to be the best feature preserved, followed by manner and finally place (Owens, Benedict, and Schubert, 1972; Dubno, Dirks, and Langhofer, 1982).

2.1.6. Investigations of the roles of audibility and effective audibility

A hearing-impaired listener has to be able to hear the acoustic cue in order to identify the consonant and in addition be able to utilize the spectral information in the acoustic cue. Raz and Noffsinger (1985) compared the perception of 'burst + transition' and transition-only stimuli by normal-hearing and hearing-impaired individuals (degree of hearing loss: mild to moderately-severe). Normal-hearing listeners showed similar performance for the two groups of stimuli. Identification of stop consonants was overall poorer for the hearing-impaired individuals. They also made more errors on the 'transition only' stimuli than on 'burst + transition' stimuli. Measurements of the root mean square intensity of the two groups of stimuli showed that the 'burst + transition' stimuli were less intense than the transition-only stimuli. While normal-hearing individuals could utilize the cues present in the formant transition alone, hearing-impaired individuals need additional information to identify the consonants. Turner and Robb (1987) examined the effect of spectral cue audibility in recognition of stop consonants and found that even when speech was well above threshold at all frequencies, the identification of phonemes by hearing-impaired individuals was not perfect.

The relationship between audibility and speech recognition can be quantified by the Speech Intelligibility Index (SII) (ANSI, R1997). The SII assumes that speech recognition increases in direct proportion to speech spectrum audibility. It is calculated from the hearing thresholds of the listener, and the long-term average of the speech and noise reaching the ear of the listener. The SII is the sum of contributions of the proportion of the speech dynamic range that is above the listener's threshold in each third-octave band, weighted by the relative importance of the band to speech intelligibility.

Ching, Dillon and Byrne (1998) examined the extent to which this audibility model could predict the speech recognition performance for individuals with sensorineural hearing loss ranging from mild to profound degrees. Subjects were fourteen normal-hearing and 40 hearing-impaired adults with mild to profound degrees of sensorineural hearing loss. Speech recognition performance was measured using filtered Bench-Kowal-Bamford (BKB) sentences (Bench and Bamford, 1979). SII was calculated using a third-octave band method.

For the normal-hearing listeners, the SII provided a satisfactory description of the speech recognition results. For the hearing-impaired individuals, the SII did not predict the scores accurately. The performance of some subjects was better than the theoretical prediction, especially at low sensation levels, and the performance of other subjects was poorer than predicted at high sensation levels. The contribution of audibility to intelligibility at frequencies where hearing loss was severe was much reduced and sometimes came close to zero. Further increases in audibility sometimes had negative effects on speech intelligibility.

The hypothesis that amplified signals may not provide usable information for persons with severe-to-profound hearing loss, even when they are audible, is supported by current research on dead regions. A dead region is a region of the cochlea where there are no functioning inner hair cells (IHC's) and/or neurons. It can be characterized in terms of the characteristic frequencies of the IHC's bordering that region. The effect of high-frequency amplification on speech perception for subjects with high-frequency hearing loss with and without dead regions was studied by Vickers, Moore and Baer (2001). They used the TEN test (Moore et al., 2000) to identify the limits of any dead regions. The speech stimuli were vowel-consonant-vowel (VCV) nonsense syllables, using one of three vowels (/i/, /a/, and /u/) and 21 different consonants. In a baseline condition, subjects were tested using broadband stimuli with a nominal input level of 65 dB SPL. Prior to presentation, the stimuli were amplified using the frequency-gain characteristic prescribed by the Cambridge prescriptive formula (Moore and Glasberg, 1998). This prescriptive procedure is intended to give speech at 65 dB SPL the same overall loudness as it would have for a normal listener, and to make the average loudness of the speech the same for each critical band over the frequency range important for speech intelligibility (in a listener without a dead region). The stimuli for all other conditions were amplified in the same manner, and then low-pass filtered with various cutoff frequencies.

For subjects without dead regions, performance generally improved progressively with increasing cutoff frequency. This indicates that they benefited from high-frequency information. For subjects with dead regions, two patterns of performance were observed. For most subjects, performance improved with increasing cutoff frequency until the

cutoff frequency was somewhat above the estimated edge frequency of the dead region, but hardly changed with further increases. For a few subjects, performance initially improved with increasing cutoff frequency and then worsened with further increases, although the worsening was significant only for one subject.

The results have important implications for the fitting of hearing aids. In subjects with very flat losses greater than 75dB HL, and steeply sloping losses (>50dB/octave) one can suspect the presence of dead regions. Providing any gain to these regions will not result in improved speech intelligibility (Vickers, 2001).

2.1.7. Perception of Place of Articulation: Saliency of short duration spectral Cues

Speech contains both spectral and temporal cues. The poor frequency and temporal resolution of the auditory system of hearing-impaired individuals may contribute to difficulties with speech perception (Florentine & Buus, 1984). The utilization of short duration spectral cues by hearing-impaired adults was studied by Dubno, Dirks, & Schaefer (1987). They carried out an experiment looking at the identification of plosives as a function of stimulus duration. Based on the configuration of their hearing loss, the hearing-impaired individuals were divided into three groups: flat, gradually sloping and steeply sloping hearing loss. The stimuli were synthetic CV syllables with consonants /b/, /d/, and /g/, and vowels /a/, /i/, and /u/. The stimuli were truncated to be of 100, 70, 45, 30, 20 & 10 ms in duration. The test was a three-alternative-forced-choice identification task. Hearing-impaired listeners performed significantly poorer than normal-hearing controls. While normal-hearing listeners' identification performance was close to 100% when syllables were longer than 20 ms,

hearing-impaired listeners did not show such dramatic improvements. Hearing-impaired listeners' performance also improved after 20 ms but did not reach 100% for any syllable duration. The performance of listeners with steeply sloping hearing loss was significantly poorer than those with flat or gradually sloping hearing loss. They also found that performance with vowels /a/ and /u/ improved faster with cut off frequency than with /i/. Thus there seems to be an interaction between the slope of the hearing loss and perception of short duration cues. It is possible that steeply sloping hearing loss causes additional distortion in the spectral domain. From this study it is also obvious that there is no invariant cue to place perception that is independent of vowel context. There is place information in the first 40 ms of the stimulus, but hearing-impaired individuals are unable to utilize it for place perception.

2.1.8. Acoustic Cues to perception of voicing

Hearing-impaired listeners may be more reliant on temporal cues when spectral information is not available to them. Revoile, Pickett, Holden-Pitt, & Talkin (1982) compared speech recognition of plosive consonants in final position by normal hearing and hearing-impaired listeners (degree of hearing loss: moderate to severe). Natural tokens of the syllables /dæp/, /dæb/, /dæt/, /dæd/, /dæk/, & /dæg/ were modified to study the effects of vowel duration and plosive occlusion duration as temporal cues, and presence of murmur in the gap and vowel offsets as spectral cues. The cue modification consisted progressively of neutralized vowel duration, equalized occlusion duration, burst deletion, murmur deletion, vowel transition interchange, and transition deletions. The normal-hearing listener's performance for voiced consonants

was intact even after the murmurs and bursts were deleted from syllables whose vowel durations were equalized. Hearing-impaired listeners' performance decreased when the duration of the vowels in the voice-voiceless pairs were equalized, and decreased further when the murmurs were deleted from the closures. Deletion of bursts did not affect performance of either group significantly. Deleting formant transitions from stimuli affected the performance of normal-hearing listeners. The performance of hearing-impaired listeners was not further reduced, either by switching transitions between the voiced-voiceless pairs, or even deleting them.

It is thus possible that hearing-impaired listeners do not make use of these cues to identify voicing. The most perceptually salient cues for hearing-impaired listeners were the duration of the vowel preceding the consonant, and the presence or absence of the murmur in the gap. It appears therefore that hearing-impaired listeners rely more on temporal cues than on spectral cues. Another explanation for the poor reliance on short duration spectral cues like bursts and formant transitions is that upward spread of masking from the vowels affects perception of other less intense spectral cues. Murmurs are typically low in frequency and most hearing-impaired adults have normal hearing in the low frequencies and hence might have been able to use the murmur information.

In a subsequent study of the recognition of plosive consonants in syllable initial position with normal hearing and hearing-impaired listeners (degree of loss: moderate to severe, flat to gently sloping), Revoile, Pickett, Holden-Pitt, Talkin, and Brandt (1987) found that bursts were important for the perception of voiceless plosives, while bursts and transitions were important for the perception of voiced plosives. They further explored the contributions of the burst release and aspiration to the perception of

plosives. They introduced aspiration in the voiced cognates of the voiceless consonants and deleted the aspiration from the voiceless plosives, replacing it with silence for all plosives. Insertion of aspiration in the voiced consonants resulted in an increased rate of identification of the stimulus as a voiceless plosive. This provides evidence that hearing-impaired listeners are able to use aspiration as a cue to initial plosive identification. The identification of voiceless plosives was unaffected by deletion of aspiration, hence listeners were able to make use of the long VOT as a cue. Long VOT also produced incorrect perception of voiceless responses for voiced consonants due to the long releases.

The results of this study demonstrate that when VOT and vowel transition cues are in conflict, normal-hearing listeners identify consonants based on the transition cues while hearing-impaired individuals identify the consonant using the VOT cues. Hence temporal cues are more salient than dynamic spectral cues for hearing-impaired listeners.

From the above studies we can summarize that the most salient features in perception of plosives by hearing-impaired listeners are the duration cues provided by the vowel to differentiate between voiced and voiceless sounds. For place of articulation they rely more on static spectral information, rather than the information in the formant transition. They are also able to utilize short duration spectral cues in conjunction with temporal cues, as long as the cues are audible. Individuals with steeply sloping sensorineural hearing loss have more difficulty with consonant perception because the slope of their hearing loss interacts with the speech spectrum such that even some of the more robust static cues are distorted to create ambiguity in perception. Formant

transitions, which provide a sufficient number of cues for the perception of speech by normal-hearing individuals, are not well utilized by hearing-impaired individuals.

2.1.9. Perception of Fricatives

Zeng and Turner (1990) studied the acoustic cues sufficient for the recognition of voiceless fricatives by normal-hearing and hearing-impaired individuals. They conducted two experiments in order to assess the role played by audibility and supra-threshold discriminability.

In the first experiment, four normal-hearing listeners heard CV syllables with /s/, /f/, /θ/, and /ʃ/ and the vowel /i/. The two kinds of acoustic cues used to differentiate between the voiceless fricatives were the formant transitions from the fricative to the vowel and the steady state spectral shape differences between the different fricatives. There were three conditions 1) natural tokens normalized for duration and amplitude, 2) synthesized syllables, and 3) truncated fricative portions from synthesized tokens. The synthetic syllables were synthesized using a parallel resonance procedure developed by Klatt (1980), to be as close to natural tokens in acoustic characteristics as possible. Performance on synthesized syllables was compared to that for natural tokens before modifying them for condition three. Speech recognition performance was measured using a four-alternative forced-choice (4AFC) procedure. Intensity was varied from 20 dB SPL to 120 dB SPL, in 10dB steps, until 100% recognition was achieved.

The audibility spectrum for speech stimuli at fricative and transition locations was calculated. Fricative portions were 20-30 dB lower than the level of the vowel.

Transitions were 5-10 dB lower than the vowel. Results showed that while fricative portions were inaudible at low presentation levels, (e.g., 20 dB SPL), recognition was still above chance level (55%). This was due to the fact that even at such low presentation levels, the transition portions were still audible.

A recognition score of 64% was obtained using synthesized stimuli containing only transition information. Once frication was audible, the recognition score rose to 94% at 40 dB SPL. When frication alone was audible, it was sufficient for recognition. Cues above 1500 Hz were sufficient for accurate perception of frication (92%). Acoustic information below 1500 Hz did not enhance performance.

These findings indicate that normal hearing listeners can use information about formant transitions to identify fricatives at low intensities. At higher presentation levels, however, both frication and transition play an important role. At very high presentation levels, frication is a sufficient cue for recognition.

In the second experiment, three hearing-impaired adults with moderate-severe hearing loss were tested. The stimuli were the synthetic stimuli presented at levels from chance performance to 100% or up to Loudness Discomfort Levels (LDL). Hearing-impaired individuals were not able to use transition cues as efficiently as normal hearing adults. Performance was primarily due to audibility of frication, limited by their LDL's. Frication only was tested at 100 dB SPL, 110 dB SPL, and 120 dB SPL. Presentation level implies that frication is at 90 dB SPL, since the presentation level is referenced to the vowel. Hearing-impaired subjects perform as well normal-hearing subjects when the frication alone is audible.

It appears that steady state spectral shapes cueing the identity of the fricatives are available to hearing-impaired adults when the frication noise is audible. Dynamic cues in the formant transition may not be utilized by hearing-impaired individuals. This inability to use the transition while frication is inaudible, and the inability to tolerate loudness of the adjacent vowel when frication is audible, may explain poor fricative recognition scores. At Loudness Discomfort Levels for vowels, frication may still be inaudible.

Hearing-impaired adults are also able to use the information present in the vowel offsets, and vowel duration and fricative duration to differentiate between voiced and voiceless fricatives. Any one of these cues is sufficient to cue the voicing of the fricative (Revoille, Holden-Pitt and Pickett 1985). It appears that the static feature of duration is more robust than the acoustic cues present in the actual frication to distinguish between voiced and voiceless fricatives.

2.2. Performance by hearing-impaired listeners in noise

Factors like background noise, reverberation, and upward spread of masking from lower frequencies, tend to mask the soft, yet important information in the speech spectrum. Given the poor frequency resolution of the impaired ear, the detail in the speech spectrum gets reduced much further. Unlike normal-hearing listeners, hearing-impaired adults cannot utilize the redundancy in the speech signal under degraded listening environments. Static spectral cues, such as vowel duration, and spectral envelope, are more relevant for consonant recognition to the hearing-impaired individual than dynamic cues like formant transitions. As a result, speech recognition performance of hearing-impaired adults tend to fall below that of normal hearing adults even if they

only have a mild hearing loss (Schum et al., 1991, Needleman and Crandell, 1995, Bochner et al, 2003). This decrement in performance cannot be explained by reduced audibility alone. Reduced spectral and temporal information also plays a role in the poor speech recognition performance in noise by hearing-impaired individuals.

Baer, Moore and Kluk (2002) investigated the ability of persons with hearing loss to use the high frequency information in speech masked by noise. Ten subjects with high-frequency hearing loss, including eight from the study of Vickers et al (2001), were tested. Five subjects had dead regions on the cochlea (as determined using the TEN test) starting below 2 kHz, and five had no dead regions. Speech stimuli (VCV) at a nominal level of 65 dB were mixed with spectrally matched noise, amplified according to the Cambridge prescriptive formula, and then low pass filtered. The noise level was chosen separately for each subject to give a moderate reduction in intelligibility relative to listening in quiet.

For subjects without dead regions, performance generally improved with increasing cutoff frequency up to 7.5 kHz, on average more so in noise than in quiet. For most subjects with dead regions, performance improved with cutoff frequency up to 1.5-2 times the edge frequency of the dead region, but hardly changed with further increases. Calculations of speech audibility using a modified version of the articulation index showed that making high-frequency components of the speech audible for subjects with dead regions often failed to provide the benefit of increased speech audibility of the high frequencies.

Additional evidence that amplification can provide benefit for some persons with hearing loss, even in noise comes from a recent study by Turner and Henry (2003).

Speech was presented in a background of multitalker babble to listeners with various severities of hearing loss. The signal was low-pass filtered progressively at different cutoff frequencies and speech recognition was measured as additional high-frequency speech information was provided to the hearing-impaired listeners. Results show that, regardless of hearing loss or frequency range, providing audible speech resulted in an increase in recognition score. However, the gains in speech recognition were small, and the maximum score obtained by listeners was low, in part due to background noise.

It is therefore possible to provide cues to improve speech recognition in quiet and in noise and more research is necessary to determine just how one can achieve improved speech intelligibility.

2.3. Signal processing to improve speech recognition of hearing-impaired listeners

The traditional signal processing approach for hearing aids has been frequency-specific amplification. As was discussed in earlier sections, spectral information even if audible, may not be usable by the hearing-impaired person. Thus it is important to investigate other signal processing techniques that may help to address deficits in frequency resolution.

Deficits in frequency resolution are thought to be caused by broadened auditory filters in the cochleas of persons with sensory hearing loss (Moore and Glasberg, 1986). Broadened auditory filters would have a negative effect on the perception of both spectral and temporal cues in the speech signal. Studies in which the speech spectrum has been smeared to simulate the loss of intelligibility resulting from poor frequency resolution have demonstrated a reduction in speech recognition performance in normal hearing listeners (Boothroyd et al 1996, Baer and Moore, 1993).

2.3.1. Spectral Sharpening

Spectral sharpening is a signal processing strategy designed to address the consequences of broad auditory filters. Several studies have been carried out to determine the efficacy of this processing approach. Bunnell (1990) used a continuously adapting filter along with different contrast weights to enhance the peaks relative to the valleys in the speech spectrum. Ten subjects with moderate to severe sensorineural hearing loss were tested. The consonants used were voiced plosives in association with vowels /a/, /i/, and /u/. Prior to the tests, listeners rated different unprocessed tokens of the same syllables on clarity as being POOR, MODERATE or GOOD.

Different contrast weights were used to provide different amounts of contrast within the speech signal. A contrast of 1 meant no difference, <1 meant a reduced contrast and >1 meant an enhancement. The contrasts were specified for the frequencies in the target spectral envelope, such that maximum enhancement was applied in the mid-frequency region. Bunnell showed a 10% increase in intelligibility scores as the contrast was increased from the reduced to the normal to the enhanced condition. The greatest amount of increase in intelligibility was seen for tokens that were rated by listeners as having moderate clarity before the tests.

Simpson, Moore & Glasberg, (1990) also investigated the enhancement of the spectral contrast between peaks and valleys as a method to improve intelligibility. They smoothed the spectrum of the speech prior to enhancement in order to remove components that could not be resolved by a normal ear. After performing a 256 point fast Fourier Transform, they calculated the auditory excitation pattern (Moore and Glasberg, 1983) and enhanced it with the difference-of-Gaussians (DOG) or 'Mexican Hat' function. This can be explained as taking a second order derivative of the speech spectrum. Centering this formula on each of the 128 magnitude values results in a signal where the peaks are amplified and the valleys are reduced resulting in a large peak-to-valley ratio.

Nine subjects with sensorineural hearing loss were tested using the BKB sentence lists and the AB word lists (Boothroyd, 1968). Subjects had to repeat the words they heard. Processing resulted in mean increases of 11.4% in identification of the sentence material and of 6.4% for the words. For initial consonants, the processing was effective for plosives and glides as well as for /f/ and /θ/. For final consonants, enhancement was

most effective for /t/ and /k/. The enhancement did cause some ambiguities in identification of plosives due to louder and more enhanced bursts.

Using similar processing as in the previous experiment Baer, Moore and Gatehouse (1993), evaluated nine processing schemes (three DOG values, 0.5, 1.0 or 2.0 times the ERB, along with three levels of enhancement, 0.3, 0.6, and 0.9).

They tested 11 subjects with mild to severe sloping sensorineural hearing loss. Subjects wore a linear hearing aid set at MCL. The Adaptive Sentence List (MacLeod and Summerfield, 1990) test was used. While speech recognition scores failed to show improvement with enhancement, subjective ratings indicated that subjects preferred the quality of moderately enhanced speech.

In a separate experiment, Baer, Moore, and Gatehouse (1993), also investigated the effect of different amounts of fast-acting amplitude compression along with the spectral enhancement. The enhancement strategy used for this experiment was a function that was the difference between two excitation patterns. The gain function was modified to limit the maximum gain to 20 dB at any frequency, keeping the gain value within a region of comfort. The enhancement was applied to the original magnitude spectrum, resulting in the enhancement of the major spectral features. The fine-grained spectral detail was not smoothed by this enhancement as in the other conditions. The same stimuli were used as in the previous condition and the same subjects as in the previous condition were tested. Moderate compression with moderate spectral contrast gave the highest intelligibility ratings. This finding was also supported by statistically significant response time measures on a sentence verification task.

Thus it appears that enhancing parts of the spectrum in relation to the others does give rise to moderate increases in speech intelligibility. It is important to investigate the specific regions of enhancement that did improve the intelligibility scores so as to improve upon the technique.

2.3.2. The use of sine waves to code speech

Speech intelligibility is preserved when speech is modeled with as little as three sine waves (Remez et al., 1981). Linear Predictive Coding (LPC) was used to derive sinusoids every 15 ms from the changing center frequencies and amplitudes of a naturally produced sentence. This type of coding will track the formants in the original signal. Normal hearing listeners identified the sentence correctly when all three sinusoids were present. Even in the absence of traditional cues to phonetic identity, like formant transitions and short-term acoustic spectra, phonetic perception was fairly intact.

Remez and Rubin (1990) further demonstrated that normal-hearing listeners rely primarily on the time varying frequency information and gross spectral amplitude changes to perceive the phonetic structure. They had normal-hearing listeners transcribe the number of syllables heard in sinusoidal replicas of speech of the type synthesized by Remez (1981). They tested four different conditions. In the first condition, the frequency and amplitude variation were preserved. In the second the frequency variation was preserved, but the energy level was constant. The third condition had misleading amplitude contours imposed on the frequency variation. The fourth condition had natural amplitude variation, but constant frequency. The hypothesis was that if phonetic coding were preserved in the time-varying properties of frequency variation, then transcription

accuracy would be good for all the conditions except the fourth. The results supported the hypothesis. Subjects accurately identified the number of syllables in the first three conditions, but were not able to perceive phonetic information in the fourth condition. This finding suggests that listeners are sensitive to the gross changes in spectral shape and not to the finer detail. Fine variation in amplitude does not affect perception greatly, as long as the frequency variation in the signal is present. Despite the absence of the original speech signal, information about the variation in vocal resonances is preserved in the frequency carriers of sine wave substitutes and can give rise to phonetic perception.

This study supports evidence from studies using sine wave modeling of speech (described later), where speech intelligibility is preserved when speech is modeled with a limited number of sine waves. In sine wave modeled speech, the gross spectral variation in the spectral envelope is preserved although the fine variation in envelope structure is not retained.

2.4. Evidence of improved speech intelligibility with reduced spectral information:

Application to cochlear implants

The function of the speech processor in a cochlear implant is to convert speech information into electrical stimuli. A major thrust in cochlear implant research focuses on how to deliver the maximum amount of speech information with the fewest number of electrodes.

One of the most popular cochlear implant processors, the Nucleus multi-electrode intra-cochlear implant, used a feature extraction strategy (MPEAK) which was formant based. The strategy extracted the fundamental frequency and the first two formants along

with the amplitude information from the incoming signal. This frequency information was divided into logarithmic pass bands and used to activate electrodes in the corresponding regions of the cochlea.

In 1989, a new speech processing strategy was developed at the University of Melbourne. The incoming signal was analyzed using 16 digital bandpass filters, each of which was assigned tonotopically to one of 16 electrodes. In each cycle, the processor detected 6 channels with maximum amplitude and activated the corresponding electrodes. The performance was expected to be superior with this Spectral maxima speech coding strategy (SPEAK) because at any given time the six electrodes conveyed information about the formants and some of the spectral detail of the envelope.

In a study that compared the two systems, subjects alternately wore both processors for 2 four-week periods each and were tested at the end of each period. A battery of speech tests was administered including nonsense syllables, monosyllabic words and sentence tests in quiet. Recognition of speech in noise was tested using spondees and sentences at +5 and +10 dB signal-to-noise ratio. Subjects also had to respond to a questionnaire.

The study showed that the SPEAK strategy resulted in significantly improved performance on all tests. The results of the questionnaire also indicated that 80% of the respondents felt they heard well with the SPEAK strategy for everyday listening situations (Skinner et al, 1994).

At present there are two different strategies for speech coding that are available for cochlear implants. One of them, a fixed-channel strategy uses a relatively small

number of input channels, typically from 6-12, with the number of output channels being the same as the input channels. Shannon et al, (1995) reported near perfect (i.e. >90%) recognition of words in quiet, in simulation which used four frequency channels. This finding was validated by Dorman and Loizou (1998), who compared the performance of normal-hearing listeners with similarly processed speech to the performance by cochlear implant patients with the same number of channels. The implant patients with the best scores performed as well as normal hearing adults. For more difficult sentence material, 8-10 channels are required for maximum performance. The required number of channels for maximum levels of performance also increases with decreasing signal-to-noise-ratio (Loizou, Dorman and Tu, 1999; Dorman, Loizou, Fitzke and Tu, 1998; Fu, Shannon and Wang, 1998).

The second speech coding strategy known as a channel-picking strategy uses a large number of input channels (around 20) and a smaller number of output channels (6-9). The rationale behind using this strategy is that speech can be understood when only the peaks in the short-term spectrum are transmitted. Studies have shown maximum levels of performance for sentences processed in noise when picking 4 out of 6 channels, 6 out of 12, and 12 out of 16 channels (Loizou, Dorman, Tu & Fitzke, 2000).

Thus, although transmitting fewer channels, channel-picking strategies seem to provide the same or better scores than fixed-channel strategies (Nakisa, 2000; Dorman, Loizou, Spahr, and Maloff, 2002).

Presenting the peaks in the short-term spectrum can be advantageous especially in quiet or in the presence of steady state background noise. The peaks of

speech (which include the formants) carry information important to speech perception. This information may be coded as the gross spectral changes in the speech envelope. The strategies that have preserved these gross spectral changes in the speech envelope have been successful in relaying speech information

2.5. Sine Wave Modeling

Sine wave modeling is a parametric method for representing speech signals. It was developed in the mid- 1980's and has since found applications in a wide variety of areas such as music synthesis, speech synthesis, signal enhancement and speech and audio coding. The original technique developed by McAulay and Quatieri (1986) is a speech analysis/synthesis procedure in which the signal is reproduced using a limited number of sine wave components. The system is based on a sine wave representation of speech that incorporates a model of speech production, but which is independent of voicing state and pitch.

In the signal analysis portion of the processing, the incoming signal is divided into overlapping fixed duration segments. Each segment is then multiplied by a Hamming (raised cosine) window. The spectrum is computed using a fast Fourier Transform (FFT). The desired number of peaks is selected from the magnitude spectrum. The frequency, magnitude and phase of each peak are saved. The output signal is synthesized by generating one sinusoid for each selected peak using the original frequency, amplitude and phase values. The McAulay and Quatieri algorithm incorporates a procedure for tracking the path taken by each sine wave as a function of time (when it is "born," when it continues in the next time frame, and when it "dies"). In order to maintain the

continuity between successive analysis frames, an interpolation procedure is followed that uses the continuity information in the computation of smooth frequency transitions from frame to frame in order to minimize phase discontinuities at the frame boundaries.

The original motivation behind using a sine wave model was to code speech with as few parameters as necessary to permit efficient transmission. Representing the entire speech spectrum with a small number of sine waves has the advantage of automatically identifying formant information (due to picking the highest concentrations of energy in the speech signal), of preserving the spectral envelope that has been shown to be very important to consonant perception, and may automatically provide a means of improving signal-to noise-ratio. In addition, once sine wave modeling has been carried out, it makes it easy to implement other forms of processing such as frequency shifting, and time-base modification (Quatieri and McAulay, 1986).

While using sinusoids to represent speech, only the most intense spectral peaks are reproduced, while the less intense speech and noise components are discarded. For example, a 256-point Fast Fourier Transform results in a representation of the spectrum of the windowed speech segment at 128 equally spaced frequencies. If 16 peaks are chosen based on order of diminishing amplitude for the synthesis, the resultant signal will show an improvement in signal-to-noise ratio up to +12dB assuming that the suppressed peaks and valleys are all uncorrelated Gaussian noise with equal variances (Kates, 1994). Given the nature of the processing, it is very difficult to estimate an output signal- to-noise ratio for speech in noise. Sine wave modeling also increases the spectral contrast in the signal, since only the most intense peaks are reproduced, which usually tend to occur around the speech formants, leaving gaps in between (Kates, 1991).

The spectral sharpening effect of sine wave modeling is illustrated in next section shows three figures. Figure 1 illustrates a power density spectrum (PDS) of the vowel /i/ produced by a male speaker. Figure 2 shows the same vowel modeled with eight sine waves. Figure 3 shows the PDS of the vowel reproduced with four sine waves. As can be seen from the figures, in the regions between 2000 and 4000 Hz there is a sharp contrast between the peaks and valleys in the spectrum for the sine wave modeled vowels and the rest of the spectrum is much diminished.

Figure 1. (unprocessed /i/)

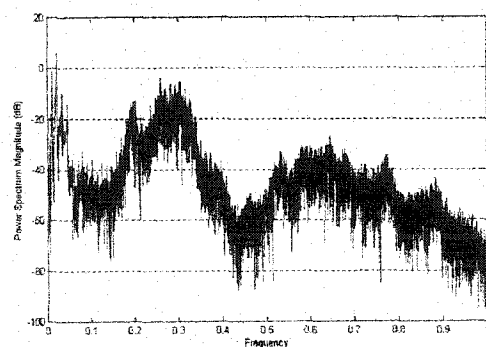


Figure 2. (/i/ with 8 sine waves)

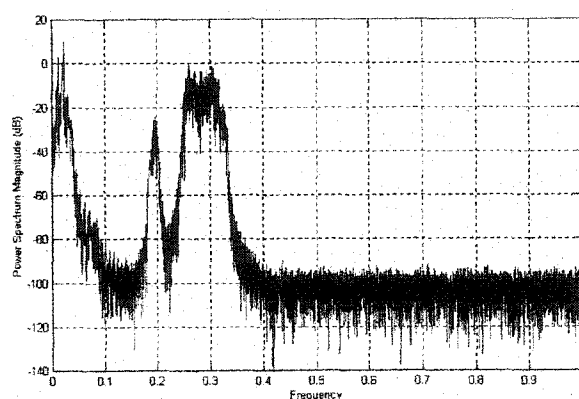
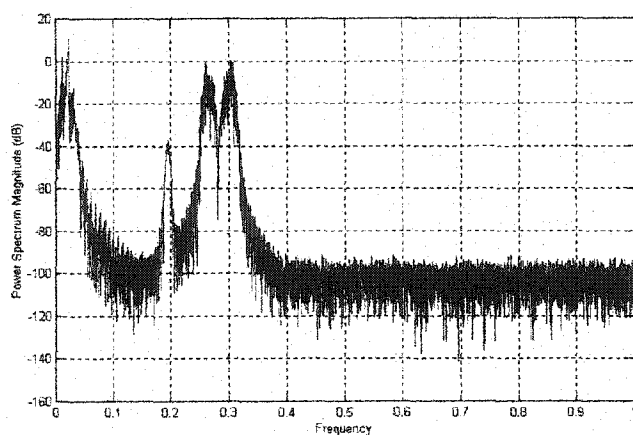


Figure 3. (/i/ with 4 sine waves)



Thus, as illustrated, sine wave modeling of speech causes an increase in the contrast between the peaks and valleys in the speech spectrum.

Using the technique proposed by McAulay and Quatieri (1986), Levitt et al (1993) investigated the effectiveness of using sine wave modeling as a technique for improving speech intelligibility in noise. Subjects were four listeners with normal hearing and two listeners with moderate to severe sensorineural hearing loss. Five sentences spoken by a male talker were processed using 4, 8, and 16 sine waves. Sentences were presented both in quiet and cafeteria noise with a signal-to-noise ratio of +5 dB. Paired comparison judgments of intelligibility were used in a round robin tournament format to determine how the number of sine waves used to model the speech would affect judged speech intelligibility.

The unprocessed condition was always judged as being the most intelligible in the quiet condition. For the signal-to-noise ratio of +5dB, sentences processed with 16 sine waves were found to be equally intelligible to the unprocessed version. For both quiet and noise conditions, as the number of sine waves decreased, the intelligibility was judged to decrease. There were no systematic differences in the performance of normal-hearing and hearing-impaired subjects.

Kates (1994) investigated the effectiveness of sine wave modeling as a single microphone technique for enhancing speech intelligibility in noise. He tested five normal-hearing adults on consonant recognition and perceived intelligibility. The stimuli used were isolated VC and CV tokens read by a male speaker. Thirteen consonants were used: /p/, /t/, /k/, /b/, /d/, /g/, /f/, /θ/, /s/, /ʃ/, /ð/ and /z/. The vowel contexts were consonant+/a/, /a/+consonant, and /i/+ consonant. The tokens were presented at +25 dB

or +5 dB signal-to-babble ratios, and were denoted as quiet and noise respectively. The consonants were processed to create tokens having 16 and 8 sine waves. Tokens with 16 sine waves employed peaks picked directly from the amplitude spectrum. Tokens with 8 sine waves employed peaks picked after the spectrum of the signal had been smoothed by convolution with a critical band filter shape. The task of the listeners for the consonant recognition condition was to identify the consonant heard from a randomized list of tokens. Paired-comparison judgments of intelligibility were also obtained.

Sine wave modeling reduced the accuracy of consonant recognition in quiet and in noise. A reduction in the number of sine waves resulted in a transmission loss for the speech features of place and frication. The weak fricatives were identified with the least accuracy, while the sibilants and the velar plosive was identified with the greatest accuracy. Manner features such as voicing and duration seemed to be better preserved.

Kates postulated that the perceptual correlate of sine wave modeling is that of adding noise to the speech signal. In the perceived intelligibility judgments, the sine wave speech was never considered to be more intelligible than normal speech. Intelligibility of the speech modeled with 8 sine waves was judged to be poorer than with 16 sine waves under all test conditions. These results did not provide evidence that sine wave modeling improved speech recognition performance.

One of the most important variables that can affect the quality of the speech while synthesizing waveforms is the method of concatenating the windowed structures. Kates (1994) used linear frequency interpolation between adjacent segment peaks and a strategy for avoiding phase discontinuities across frame boundaries. After Kates published his findings, Balachandran and Levitt (1997) used the Overlap-Add procedure

(Allen, 1977) for concatenating windows and found a great improvement in sound quality. Subjective reports of listeners who compared sine wave speech processed using the overlap-add method and the linear interpolation method indicated that the overlap-add method resulted in fewer perceptual distortions than linear interpolation. This might explain why sine wave modeled speech in Kates' study was never judged to be as intelligible as natural speech. It might also have affected the consonant recognition performance of the subjects.

While Kates' study did not show any improvement in speech intelligibility for normal hearing listeners, it was felt that with the new implementation of sine wave modeling different results might be obtained with both normal-hearing and hearing-impaired listeners.

2.6. Report of Early Studies with the Overlap-Add technique

The focus of the initial experiments was to determine the point at which listeners could no longer tell the difference between natural speech and sine wave modeled speech. Also of interest was to determine the minimum number of sine waves necessary for consonant recognition. The studies are described in more detail below.

2.6.1. Consonant Recognition

Balachandran and Levitt (1997) carried out a study to determine the number of sine waves necessary for the identification of individual consonants in a nonsense syllable context. Four normal-hearing listeners were tested. The number of sine waves was varied adaptively. The stimuli were 48 VCV syllables of 16 consonants (/p/, /t/, /k/,

/f/, /θ/, /s/, /ʃ/, /b/, /d/, /g/, /v/, /ð/, /z/, /n/, /ŋ/, & /m/ in three vowel contexts /a/, /i/, & /u/).

The tokens were presented individually and subjects had to push a button corresponding to the consonant heard. The number of sine waves was varied from 2 to 42 in steps of two.

Results revealed that subjects needed approximately 6 sine waves to accurately recognize individual consonants. This finding is contrary to Kates' results that indicated that consonant accuracy was reduced relative to normal speech when the number of sine waves was either 8 or 16. The results provide evidence for the superiority of the overlap-add procedure as part of implementing sine wave model processing.

2.6.2. Perceptual similarity

The next set of experiments was done to determine the minimum number of sine waves required to make natural speech and sine wave modeled speech indistinguishable to listeners. An adaptive procedure was used with the oddity paradigm. In this paradigm, listeners were presented with three tokens. The subject's task was to select the token that sounded different from the other two. The odd token could be the natural token or the sine wave synthesized token. If subjects picked the correct token twice in a row, the number of sine waves presented was increased. If the response was incorrect, the number of sine waves decreased. Thus the procedure converged at the 70.7% level of performance (Levitt, 1971). For sine wave speech to be indistinguishable from natural speech only 30 components were needed in quiet and approximately 15 sine waves in the presence of background noise (signal to noise ratio of +15 dB).

The experiment was repeated with three hearing-impaired adults with moderate to severe sensorineural hearing loss. Three amplification conditions were included: a flat frequency response, selective amplification applied after sine wave model processing (NAL-R prescriptive procedure), and selective amplification (NAL-R procedure) prior to peak selection for sine wave model processing. All subjects were tested in quiet. The number of sine waves required for the same task by hearing-impaired subjects was approximately half of that required by normal-hearing listeners. The improvements observed with frequency shaping were confounded with substantial learning effects seen over experimental trials. The group of subjects tested was small and inter-subject variability was high.

The goal of the initial studies with hearing-impaired listeners was to determine the minimum number of sine waves needed to make sine wave synthesized speech indistinguishable from natural speech tokens. It was found that performance did not stabilize, but rather continued to improve throughout the experiment. This outcome could be attributed to the presence of processing noise. There is a very low level noise in the background (probably caused by discontinuities in frequency and in phase between successive time frames) that was identified by the subjects when listening for the odd token in the oddity paradigm. Verbal reports by subjects indicated that while the speech quality of the processed tokens was excellent, the presence of this processing noise provided them with a cue to the identity of the odd token in a set. Hence adding more sine waves in a window up to a limit of 40 did not change the perceived difference between the tokens. Subjects consequently performed near the ceiling by using the

processing noise as a cue. From the results of the above experiments and subjective reports of the subjects it appears that the point of no difference between natural speech tokens and sine wave tokens is somewhere between 20 and 40 sine waves.

2.7. Rationale

The evidence from the pilot studies indicates that normal hearing adults can recognize individual consonants with 8-12 sine waves. It is expected that hearing-impaired individuals would also need similar or perhaps a fewer number of sine waves for consonant recognition. This hypothesis stems from the fact that as more sine waves are added they are lower in amplitude and a hearing-impaired listener may not be able to effectively use the information encoded by the lower amplitude cues.

Sine wave modeling may have the benefits of coding the signal in a way that will allow hearing-impaired listeners to obtain more information from the processed speech signal than they would from an unprocessed signal due to spectral sharpening effects, effective presentation of speech peak information, and preservation of important temporal envelope information.

The use of sine wave modeling in the presence of noise is expected to improve the signal-to-noise ratio. When using sine wave modeling there is, of course, a danger that peaks in the noise will be chosen instead of speech peaks. However, the peaks of speech in any short time interval will occur more consistently in the same position in overlapping time windows and thus have a greater likelihood of being picked by the algorithm over the peaks of noise which due their random nature may not occur in the same position in each FFT window. Once the highest peaks are selected, the rest of the

signal containing both speech and noise are discarded. If the majority of the peaks chosen are speech, the processed signal should have a better signal-to-noise ratio.

The present study was designed to answer the following questions about the recognition of sine wave modeled speech:

- 1) What is the effect of varying the number of sine waves on consonant recognition of speech in quiet?
- 2) What is the effect of varying the number of sine waves on consonant recognition of speech modeled in noise (as a function of signal-to-noise ratio)?
- 3) Are there differences in recognition of sine wave modeled speech by normal hearing and hearing-impaired listeners?
- 4) Are the most intelligible sine wave synthesized speech tokens more intelligible than natural speech tokens?

CHAPTER 3

Methods

3.1. Participants

Twenty hearing-impaired individuals participated in this study. The hearing-impaired adults ranged from 30-78 years in age. Participants had bilateral sloping (<20 dB/octave) moderate to moderately-severe/severe sensorineural hearing loss (see Table 1). Of the twenty participants recruited, five participants were female and 15 were male. All participants were binaural hearing aid users who had been wearing hearing aids for over five years and were native speakers of British English. Their optimal discrimination scores (ODS) ranged from 57-100% (Mean = 86.4%; S.D. = 14.3) as measured with the AB word list (Boothroyd, 1968) at comfortable listening levels. All participants had sensorineural hearing loss with air-bone gaps no greater than 10dB. The Olsen-Noffsinger version of the Tone Decay Test was performed at 2000 Hz (Olsen and Noffsinger, 1974). Participants had no middle ear dysfunction and showed no signs of retrocochlear pathology on tone decay and reflex decay tests.

Six normal hearing adults were recruited from a pool of graduate students at University College London. Two males and four females between the ages of 22 and 36 years served as controls. Each had pure tone thresholds better than 15dB HL at all frequencies between 250 and 8000 Hz, no clinically significant tone decay and a negative history of otologic disease. All participants were compensated for their time and travel.

TABLE 1.

Subject characteristics and audiometric data in the test ear for the hearing-impaired adults.

No	Sex	Ear	Age	Pure Tone Thresholds (dB HL)										ODS	MCL
				.25	.5	.75	1	1.5	2	3	4	6	8	%	(dB HL)
1	M	R	63	35	50	50	60	50	50	50	45	55	70	100	82
2	M	R	58	55	65	65	65	65	60	70	80	85	90	83	88
3	M	L	55	30	30	40	40	45	45	55	65	60	55	93	88
4	M	L	61	35	45	50	60	65	75	75	75	80	90	63	88
5	M	L	33	45	55	55	60	60	55	60	60	70	80	67	88
6	F	L	60	55	55	55	55	55	55	50	40	45	50	98	86
7	M	R	70	35	40	40	40	40	40	55	60	75	80	96	86
8	M	L	35	30	45	50	60	60	65	70	70	70	65	83	90
9	M	R	62	35	40	40	40	40	40	45	50	55	55	100	84
10	M	R	65	60	55	55	50	55	60	65	60	60	70	57	88
11	F	L	74	45	60	60	55	50	50	50	45	45	45	100	86
12	M	L	75	60	55	50	45	40	35	45	55	60	80	100	88
13	M	R	73	25	30	40	45	50	55	55	50	70	80	70	82
14	F	R	32	45	50	60	60	60	55	55	50	45	45	93	84
15	M	L	67	35	40	45	45	50	55	60	70	70	65	97	84
16	M	R	65	25	35	45	50	50	50	50	50	55	65	63	88
17	F	R	30	35	50	55	65	70	70	75	90	90	90	93	86
18	F	R	50	20	20	20	35	40	45	50	55	65	65	100	84
19	M	L	78	35	30	40	60	70	75	85	80	90	90	93	84
20	M	L	70	15	25	25	30	35	45	50	55	60	70	80	84
Mean			59	38	44	47	51	53	54	59	60	65	70	86	86

TABLE 2.Subject characteristics and audiometric data in the test ear for the normal-hearing adults.

No	Sex	Ear	Age	Pure Tone Thresholds (dB HL)										ODS %	MCL (dB HL)
				.25	.5	.75	1	1.5	2	3	4	6	8		
1	F	R	24	15	10	10	10	5	5	5	5	5	5	100	68
2	F	R	23	5	5	5	5	5	0	0	5	0	0	100	66
3	M	L	36	10	15	10	10	10	10	10	10	15	15	100	74
4	F	L	22	0	5	5	10	10	5	5	0	0	0	100	62
5	M	R	32	10	10	5	5	5	0	5	5	0	0	100	68
6	F	R	22	10	10	10	10	10	10	10	10	10	10	100	68
Mean			26	8	9	8	8	8	5	6	6	5	5	100	68

3.2. Stimuli

The stimuli were 20 intervocalic consonants (VCV) syllables (/p/, /t/, /k/, /b/, /d/, /g/, /f/, /θ/, /s/, /ʃ/, /v/, /z/, /tʃ/, /dʒ/, /j/, /w/, /r/, /l/, /m/, /n/) in the vowel environment of /a/, /i/, and /u/. VCV syllables have two vowel-related cues to the identity of the consonant, namely the formant transition into and out of the consonant, in addition to the steady state spectral cues in the consonant. This redundancy is typical of that available in continuous discourse and for this reason VCV's were chosen over VC or CV syllables for the study. The stimuli were selected from digitally stored pre-recorded VCV test material which had been digitized using a sampling rate of 22050 Hz and 16 bit quantization (Rosen, Faulkner, and Wilkinson, 1999). The VCV tokens recorded by the male speaker were selected. There were four different tokens of each VCV syllable yielding a total of 240 syllables (20 consonants x three vowels x four tokens). A carrier phrase "Please mark" was recorded in an anechoic chamber by the same speaker using a Bruel& Kjaer (B&K) 2231 sound level meter fitted with a 4165 microphone cartridge. The AC output of the sound level meter was recorded on a Sony 100ES digital audio tape recorder. The level of recordings was monitored on a Fostek 6301B speaker amplifier. The carrier phrase was later added to the individual VCV's at the same RMS level as the tokens.

3.2.1. Sinusoidal Modeling of Stimuli

Test stimuli were created using an in-house software program implementing sinusoidal modeling using MATLAB software (version 5.3, Mathworks, 1999). The syllables were processed to create versions of each syllable containing 4, 8, 12, and 16 sinusoids per window. The processing of the stimuli was carried out as follows.

The tokens were low pass filtered at 6 kHz using a 4th order Butterworth filter prior to processing. Successive 512 point FFTs were carried out on the stimuli. In each window position, 256 discrete spectrum samples were generated and from each magnitude spectrum the specified numbers of peaks were selected in the order of decreasing amplitude. To be selected as a peak the magnitude at a discrete frequency had to be flanked on either side by magnitudes of lower amplitude. No rule was implemented for the minimum difference between the selected peak and its adjacent peaks. The waveform was then reconstructed by inversely transforming the selected peak magnitudes and phases back into a time-domain waveform. An overlap/add procedure (Allen, 1977) was chosen to ensure perceptual smoothness in the signal. An overlap of 25% was achieved using a moving window that advanced by 75% of its width at each step.

For tokens that were processed with noise, cafeteria noise was added to the utterances prior to processing. Figure 4 shows the power density spectrum of the cafeteria noise used. The noise was added 50 ms before the speech and extended to 50 ms after the utterance. Tokens were created in quiet and with signal-to-noise ratios of +10 dB and +5 dB. The RMS of all stimuli was set equal to the average RMS of the vowel in their respective vowel groups.

Figure 4. Short-term Spectrum of Cafeteria Noise

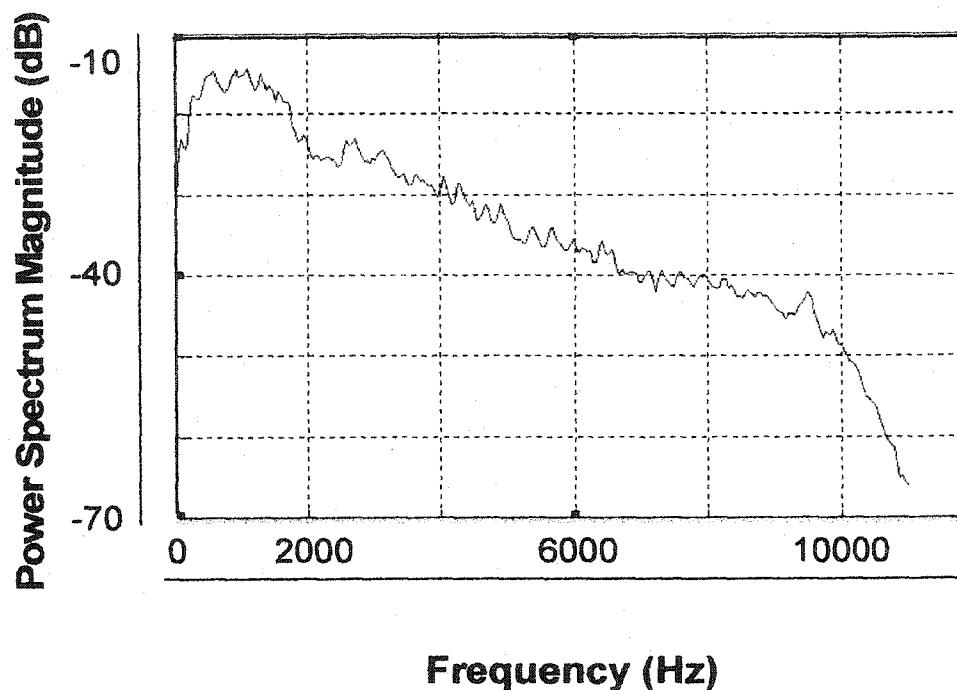


Figure 5 shows the power density spectrum of the syllable /aba/ which has been low pass filtered at 6000 Hz. Figure 6 illustrates the power density spectrum of the same syllable with 4 sine waves. As shown, the majority of information present is concentrated below 1500 Hz. When the numbers of sine waves are increased to 12 the spectral content increases to approximately 2000 Hz and with an additional peak at approximately 4000 Hz.

Figure 5. Short-term Spectrum of syllable /aba/ low pass filtered at 6000 Hz

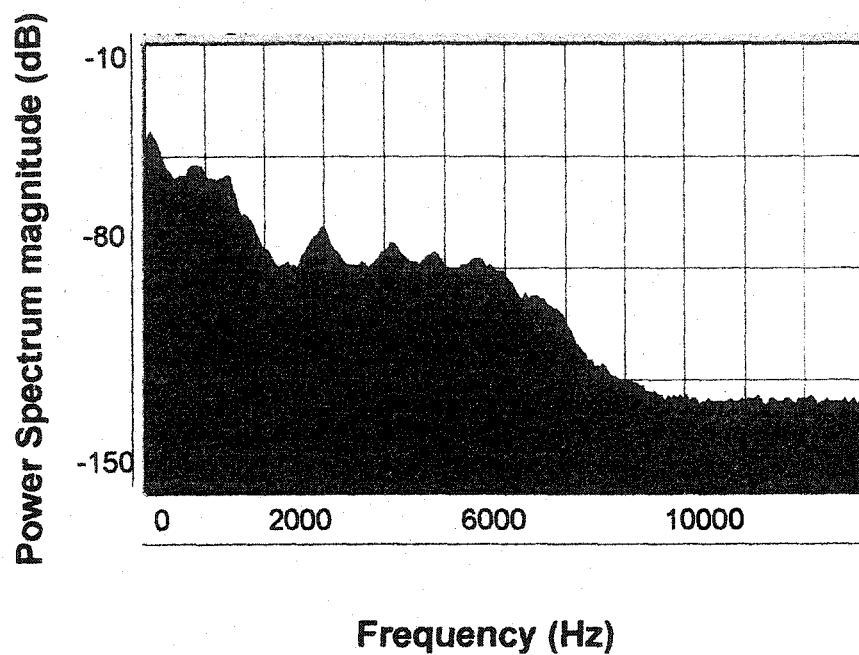


Figure 6. Short-term Spectrum for syllable /aba/ with 4 sine waves

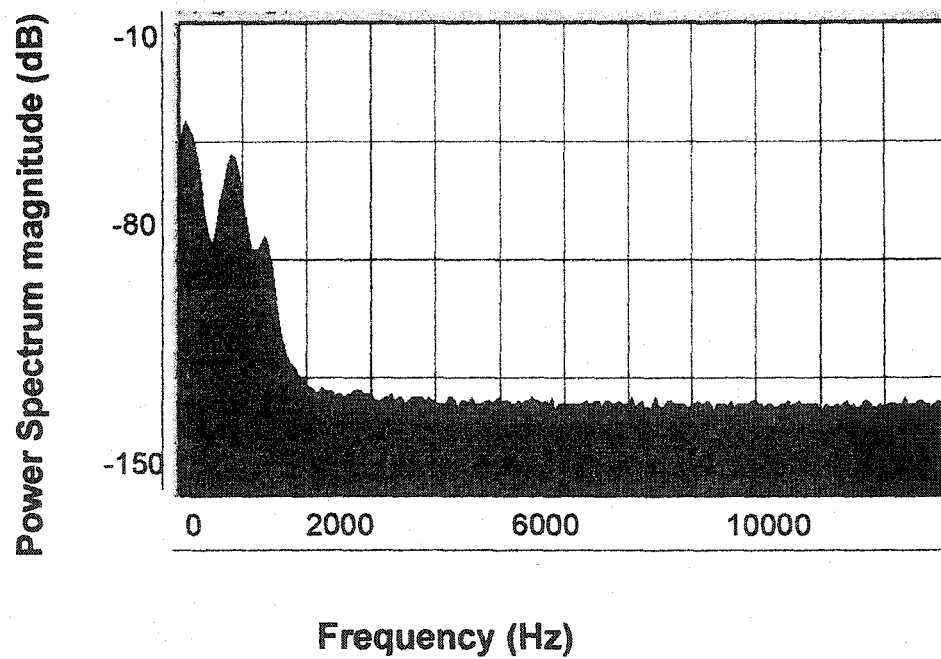


Figure 7. Short-term Spectrum of /aba/ with 12 sine waves

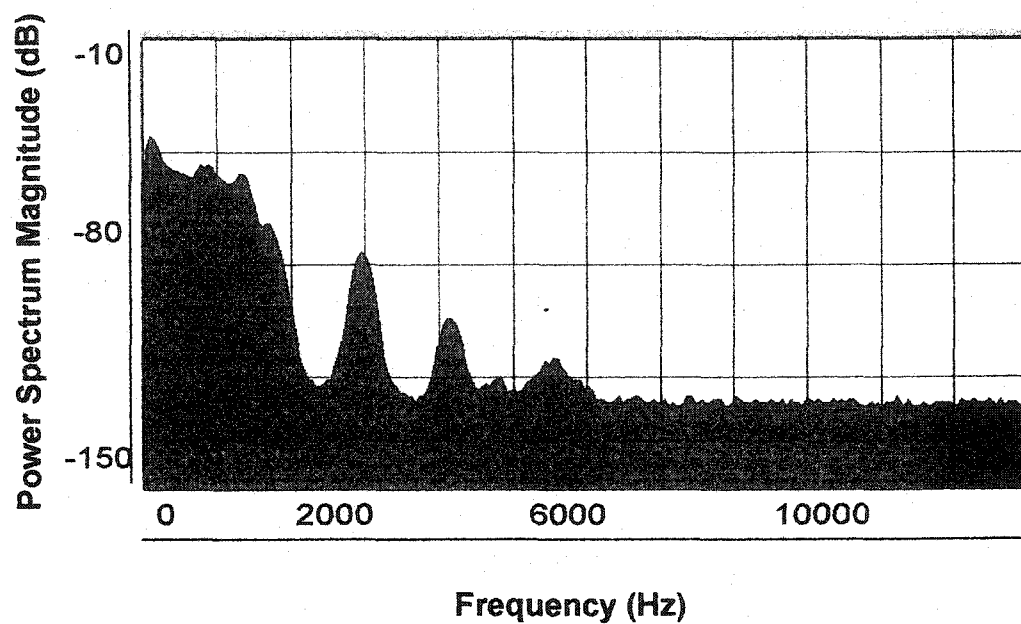
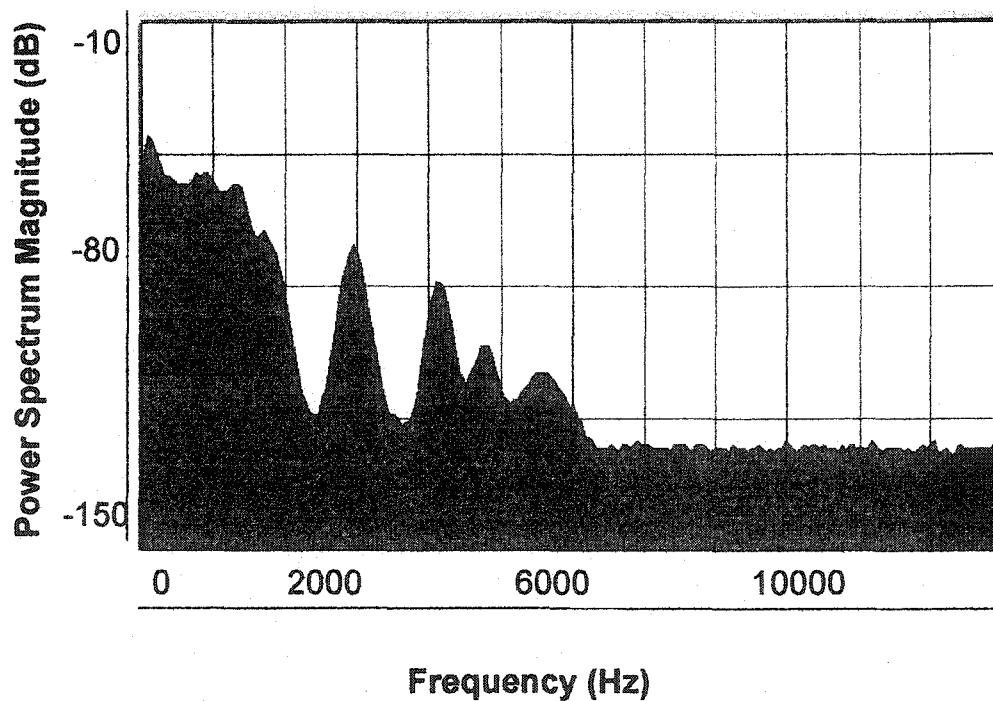


Figure 8. Short-term Spectrum of /aba/ with 16 sine waves



As shown in Figures 7 and 8, with the addition of 12 or 16 sine waves, higher frequency information becomes more available.

3.3. Instrumentation

Audiological evaluation was carried out using a Kamplex AC35 dual channel audiometer. Speech signals were played through a TEAC 80 digital audio tape recorder routed through the audiometer. All air conduction testing was performed using TDH-39 earphones with MX-41/AR cushions. Immittance testing was carried out using a GSI-37 middle ear analyzer.

The experimental stimuli were stored digitally (22,050 sampling rate, 16 bit resolution) on the hard disk of a Pentium II personal computer. The output of the digital-to-analog converter of the test computer was amplified by a Yamaha amplifier (Professional Series, Model No: P2100) with a built in attenuator with fixed 2dB steps. The 2dB attenuation was verified during calibration. The output of the amplifier was sent to a manual attenuator with 20 dB steps. The manual attenuator also had a resistor to match the impedance of the headphones and a switch selector for right or left ear. The output of the attenuator was delivered through Beyerdynamic DT48 circumaural headphones. Subject responses were made on the keyboard of the computer placed in the sound-treated booth with the subject.

Figures 9 and 10 show the frequency response of the Beyerdynamic DT-48 headphones with circumaural headphones for the right and left ears respectively. Frequency response was measured by playing a white noise signal generated by a signal

generator and analyzer (Ono Sokki Portable dual channel FFT analyzer). The output of the signal generator was passed through the headphones. The headphones were coupled to the B&K 4153 artificial ear with a B&K half inch microphone. The output was analyzed by the Ono Sokki dual channel FFT analyzer.

Figure 9. Frequency response of right earphone of Beyerdynamic DT-48 headphones under 6 cc coupler for white noise

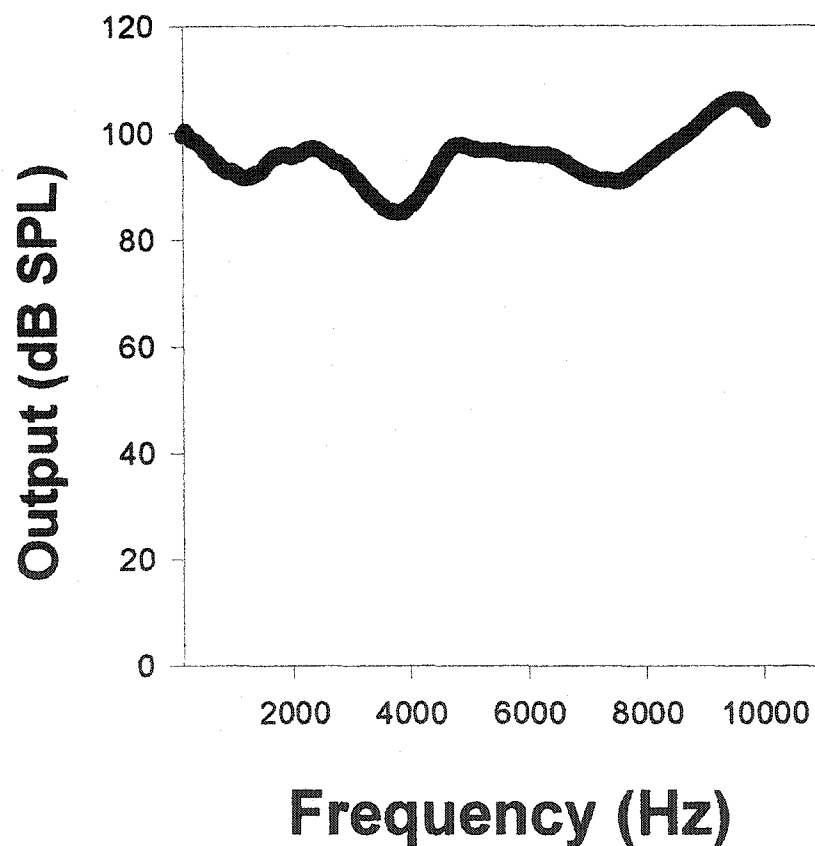
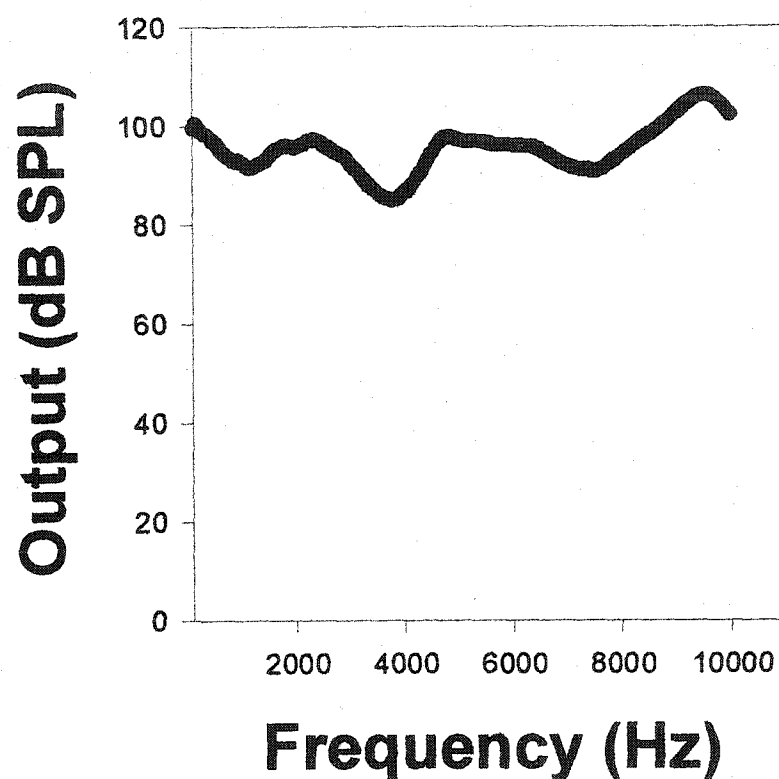


Figure 10. Frequency response of left earphone of Beyerdynamic DT-48 Headphones
under 6 cc coupler for white noise



3.4. Procedure

All testing was done in a sound-treated environment. Sound field measurements confirmed the ambient noise to be within the ANSI (ANSI S3.1-1999) standards for audiological testing.

Pure tone air and bone conduction thresholds were obtained for test frequencies 250 through 8000 Hz using ASHA guidelines for Manual Pure Tone Thresholds Audiometry (ASHA, 1977). Word recognition testing was measured using a tape

recorded version of the AB (Boothroyd, 1968) word list. Participants had to repeat the word that they heard. A performance intensity (PI) function was obtained for each subject in 10 dB increments starting at 30 dB SL. None of the scores for the participants showed any rollover, even at levels approaching Loudness Discomfort Level (LDL). Word recognition scores for each ear were within 10% of each other for both ears for all participants. The test ear was selected at random for each subject.

Before beginning the experiment the Most Comfortable Level (MCL) was measured under Beyerdynamic DT-48 headphones using the speech material to be used for the experiment. Participants were seated in a sound-treated booth, in front of a 17" VGA color monitor. The speech materials were played at 30 dB SL (re: the pure tone average of thresholds at 500 Hz, 1k Hz, and 2 kHz). A simple up-down procedure (2 dB step size) was used to determine the boundary between soft and comfortable speech (being defined as the level where speech can be heard comfortably for a long period without straining to hear). This was defined as the lower limit of the MCL (LMCL). Similarly an upper limit (UMCL) was defined as the boundary between speech that is comfortable and speech that is too loud. The MCL was obtained by averaging the LMCL and UMCL.

The test stimuli were presented at MCL using EURAUD Assessment Test Editor and Controller (ASTEC) software (Martin Brousseau, Laryngograph Ltd. and Laboratories Parole et Language [CNRS-URA_261]). This software is designed for behavioral testing of speech stimuli. Participants were seated in front of a color monitor. On the screen, they were presented with a grid of 20 buttons labeled with consonant

symbols. Participants were asked to push the button corresponding to the sound that they heard by using a mouse.

The stimuli were presented blocked by vowels. In each set, all 20 consonants x 5 stimulus versions (4, 8, 12, 16 sine waves, and unprocessed) were presented in random order. Participants were permitted to take a break after every 50 tokens. Typically, participants completed six sets of 100 tokens during each visit.

All participants completed one block of the stimuli in quiet as a practice session. Continuous feedback was given during the practice trials by displaying the correct response in a separate box near the response grid. No feedback was provided during the experimental trials.

CHAPTER 4

Results

The responses from participants were converted to proportions and an arcsine transform was applied before data analysis. Separate analyses were carried out for normal hearing and hearing-impaired individuals. Data were submitted to a repeated measures analysis of variance (ANOVA) followed by post-hoc testing (Tukey Honest Significant Difference test (HSD), Winer, 1971) to answer the following research questions.

- 1) What is the effect of varying the number of sine waves on the intelligibility of the speech tokens?
- 2) Are the most intelligible sine wave synthesized speech tokens more intelligible than unprocessed speech tokens?
- 3) What is the effect of varying the signal-to-noise ratio on speech intelligibility of sine wave tokens?
- 4) Are there differences in performance between normal hearing and hearing-impaired listeners?

Confusion matrices were generated for each sine wave and noise condition.

An information transfer analysis (Miller and Nicely, 1955) was carried out to determine how the information transferred by voicing, manner and place differed as a function of the number of sine waves used to model speech in quiet and in noise.

In addition, a spectrographic analysis of the stimuli was carried out to identify the effects of the sine wave modelling process on the stimuli. In each section the results for the normal hearing adults are presented first, followed by the results for the hearing-impaired subjects.

4.1. Experimental Results for the Normal hearing group

A summary of the analysis for the normal hearing listeners appears in Table 3. The factors in the ANOVA were: number of sine waves (S), signal-to-noise ratio (N), and vowel context (V). All three main factors were significant ($p < .001$).

Table 3.

Summary table of the repeated measures analysis of variance for the normal hearing group.

Source	Sums of Squares	dF	Mean Squares	F
Sine waves(S)	19.025	4	4.75623	151.80**
Vowels(V)	4.095	2	2.04746	67.13**
Noise(N)	39.644	2	19.822	437.25**
Persons(P)	1.030	5	0.206	13.1**
S x V	0.471	8	0.059	3.91**
S x N	1.244	8	0.155	4.99**
V x N	0.769	4	0.192	8.0**
S x V x N	0.966	16	0.060	3.83**
S x P	0.627	20	0.031	1.99*
V x P	0.305	10	0.031	1.94*
S x V x P	0.601	40	0.015	0.96
N x P	0.454	10	0.045	2.88**
S x N x P	1.247	40	0.031	1.98**
V x N x P	0.481	20	0.024	1.53
S x V x N x P	1.258	80	0.016	
TOTAL	72.216	269		

Note: * = $p < 0.05$; ** = $p < 0.01$; dF = degrees of freedom

After the analysis, the arcsine mean values were converted back to percent correct scores for ease of presentation. The mean percent correct scores for each sine wave, signal-to-noise ratio, and vowel condition are shown in Table 4. The values averaged across vowels are shown below each group in Table 4. The standard errors were very small (around 1%) and are not shown in any of the figures in this section.

Table 4.

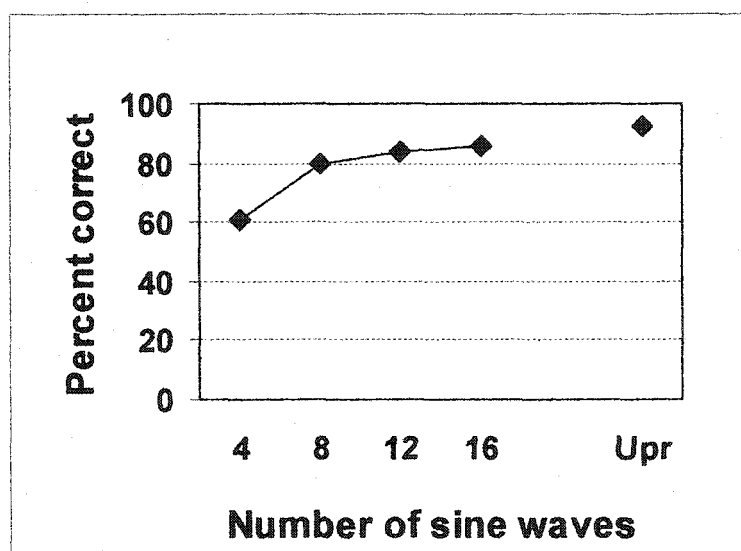
Mean consonant recognition performance as a function of number of sine waves, signal-to-noise ratio, and vowel context for the normal hearing group.

Sine waves	Quiet /a/	Quiet /i/	Quiet /u/	+10 /a/	+10 /i/	+10 /u/	0 /a/	0 /i/	0 /u/
4	80.6	85.6	84.3	69.7	58.3	53.7	38.2	35.3	29.4
Average		83.6			60.7			34.3	
8	96.9	97.0	92.7	87.1	78.5	74.7	63.0	52.9	45.3
Average		95.8			80.4			53.8	
12	98.9	96.2	93.2	92.0	84.5	78.1	68.8	66.7	47.8
Average		96.5			85.4			61.3	
16	98.3	97.0	94.4	94.7	86.4	80.1	73.8	69.3	51.6
Average		96.8			87.8			65.2	
Upr	98.9	97.2	97.6	96.5	94.6	95.2	93.4	75.8	63.6
Average		98.0			95.5			79.2	

Note: Upr = Unprocessed

The mean speech recognition performance of the normal hearing participants as a function of the number of sine waves is illustrated in Figure 11. As shown in the figure, mean speech recognition performance increased as the number of sine waves increased.

Figure 11. Mean performance of normal hearing adults as a function of sine wave condition.

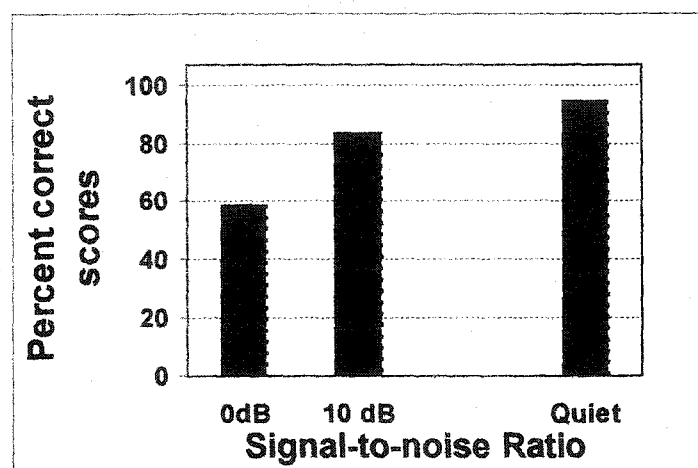


The Tukey HSD test ($<.01$ level of significance) revealed that the mean score for the unprocessed stimuli (92.5%) was significantly better than the scores for all other stimuli. The mean score for 4 sine waves (60.50%) was significantly poorer than all other conditions. The mean scores for 8 and 12 sine waves (79.43% and 83.60%, respectively) did not differ significantly, and the mean scores for 12 and 16 sine waves (83.60% and 85.60%, respectively) did not differ significantly.

Mean consonant recognition performance as a function of signal-to-noise ratio is shown in Figure 12. Inspection of the figure reveals that performance decreased as

signal-to-noise ratio decreased. Post hoc analysis ($<.01$ level of significance) reveals significant differences between all three means obtained in quiet (95%), and +10 dB SNR (84%) and at 0 dB SNR (60%).

Figure 12. Mean consonant recognition performance of normal hearing listeners as a function of signal-to-noise ratio.



The mean consonant recognition performance of normal hearing participants as a function of vowel environment appears in Figure 13. The ANOVA revealed that vowel environment had a significant effect on consonant recognition. Post hoc analysis ($<.01$ level of significance) revealed that the means for the three vowel conditions differed significantly from each other. Mean recognition of consonants in the /a/ context (86.8) was significantly better than mean performance in the /i/ context (81.3), which in turn was significantly better than consonant recognition in the /u/ context (75.1).

Figure 13. Mean consonant recognition performance of normal hearing listeners as a function of vowel context.

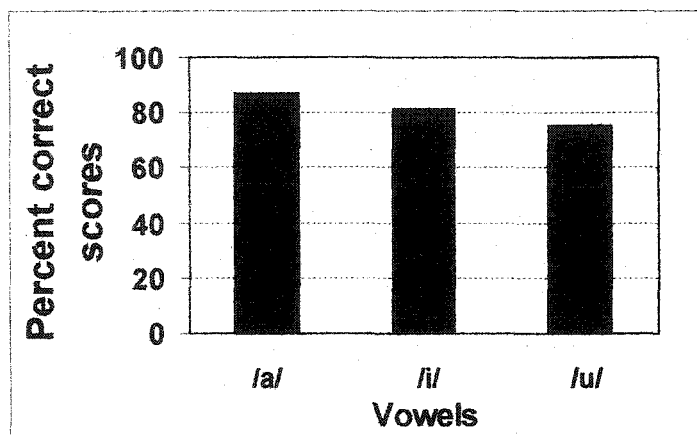


Figure 14 shows the interaction of the number of sine waves and signal-to-noise ratio. Inspection of the figure reveals that in quiet, asymptotic performance was obtained with 8 or more sine waves. In noise, the mean consonant recognition performance increased as the number of sine waves increased. However, in the noise conditions, performance on sine wave modelled speech never reached the asymptotic performance obtained with the unprocessed stimuli in that condition. Post hoc analysis ($<.01$ level of significance) revealed that the means with 4 sine waves were significantly different from each other at each signal-to noise ratio. In quiet there was no difference between the means for 8, 12, and 16 sine waves and the unprocessed condition. For the noise conditions, +10 dB SNR and 0 dB SNR, there was no difference between the means for the 8, 12, and 16 sine wave conditions. There was no difference in the means for the unprocessed conditions in Quiet and +10 dB SNR.

Figure 14. Mean consonant recognition performance of the normal hearing group as a function of the number of sine waves and the signal-to-noise ratio.

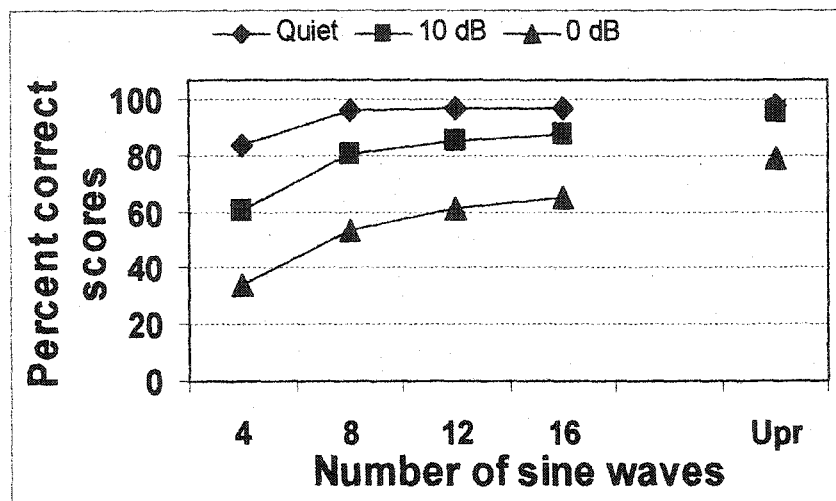
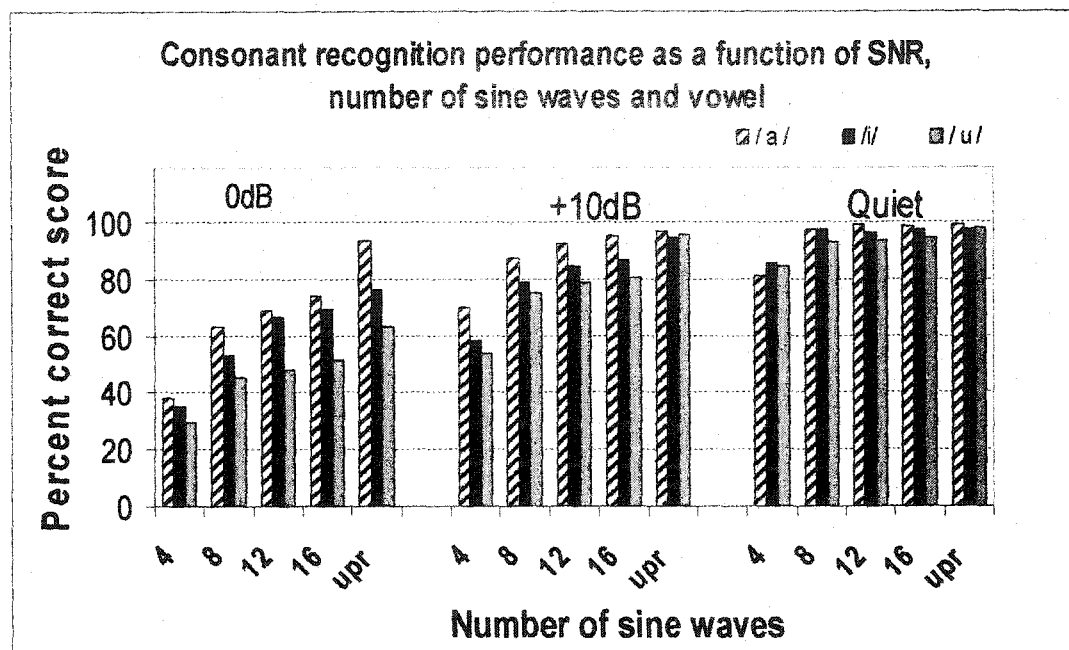


Figure 15 illustrates the significant interaction between the number of sine waves, signal-to-noise ratio, and vowel context. Overall the main findings are: In quiet, consonant recognition does not differ significantly in the 3 vowel contexts. In quiet the means of 8 and 12 sine waves and 12 and 16 were not significantly different from each other. Performance on the unprocessed stimuli was significantly better for all vowel conditions.

At +10 dB SNR, performance was significantly better for the sine wave stimuli in the /a/ context. At 0 dB SNR, consonant recognition performance was poor even for the unprocessed stimuli. Sinusoidal modelling yields degradation in all 3 vowel environments. Recognition performance was better in the /a/ and /i/ environments than in the /u/ environment.

Figure 15. Mean consonant recognition performance of normal hearing listeners as a function of signal-to-noise ratio, vowel, and sine waves.



4.2. Experimental Results for the Hearing-Impaired Group

A summary of the analysis for the hearing-impaired listeners appears in Table 5. The factors in the ANOVA were: number of sine waves (S), signal-to-noise ratio (N), and vowel context (V). All three main factors were significant ($p < .001$).

Table 5.

Summary table of the repeated measures analysis of variance for the hearing-impaired group.

Source	Sums of Squares	dF	Mean Squares	F
Sine waves(S)	16.109	4	4.027	80.392**
Vowel (V)	5.122	2	2.561	16.028**
Noise (N)	99.293	2	49.646	234.812**
Persons (P)	73.630	19	3.875	249.868**
S x V	0.737	8	0.092	3.885**
S x N	0.3112	8	0.039	2.293
V x N	0.337	4	0.084	1.510
S x V x N	0.811	16	0.051	3.268**
S x P	3.807	76	0.050	3.230**
V x P	6.072	38	0.160	10.302**
S x V x P	3.603	152	0.024	1.528**
N x P	8.034	38	0.215	13.633**
S x N x P	2.579	152	0.017	1.094
V x N x P	4.237	76	0.056	3.595**
TOTAL	229.396	899		

Note: * = $p < 0.05$; ** = $p < 0.01$; dF = degrees of freedom

After the analysis, the arcsine mean values were converted back to percent correct scores for ease of presentation. These mean percent correct scores for each sine wave, signal-to-noise ratio, and vowel condition are shown in Table 6.

Table 6.

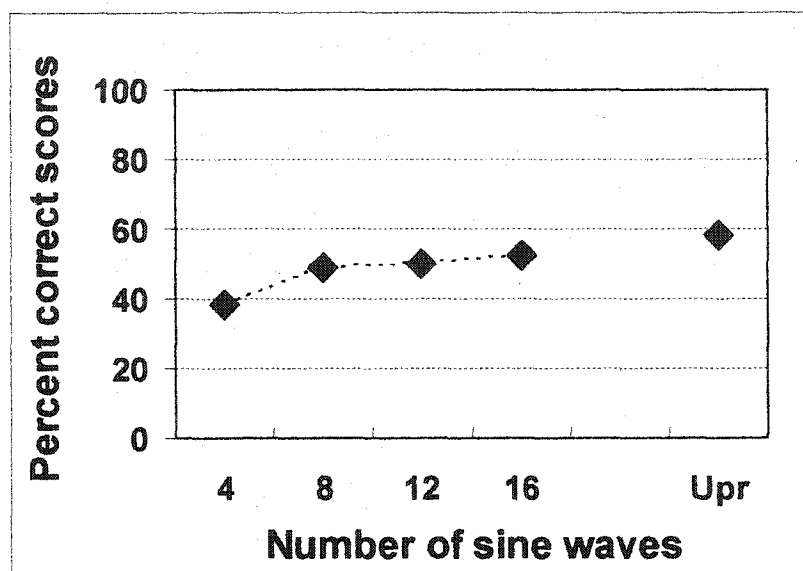
Mean consonant recognition performance for hearing-impaired adults as a function of number of sine waves, signal-to-noise ratio, and vowel context.

Sine wave	Quiet /a/	Quiet /i/	Quiet /u/	+10 /a/	+10 /i/	+10 /u/	0 /a/	0 /i/	0 /u/
4	55.5	54.8	57.7	41.1	39.0	42.4	22.8	16.6	18.1
Avg		56.1			40.9			19.2	
8	74.5	63.9	64.1	58.7	43.6	48.4	30.3	28.1	25.7
Avg		67.7			50.3			28.0	
12	73.7	65.2	64.0	62.8	49.1	50.6	31.4	26.3	25.3
Avg		67.7			54.3			27.7	
16	75.4	65.3	68.5	67.0	52.9	51.8	32.1	27.8	30.7
Avg		69.8			57.3			30.2	
Upr	79.5	70.6	72.6	69.6	58.7	59.6	42.0	34.0	34.7
Avg		74.4			62.7			36.9	

Note: Upr = unprocessed; Avg = average

The mean speech recognition performance of the hearing-impaired participants as a function of the number of sine waves appears in Figure 16. As shown in the figure, mean speech recognition performance increased as the number of sine waves increased

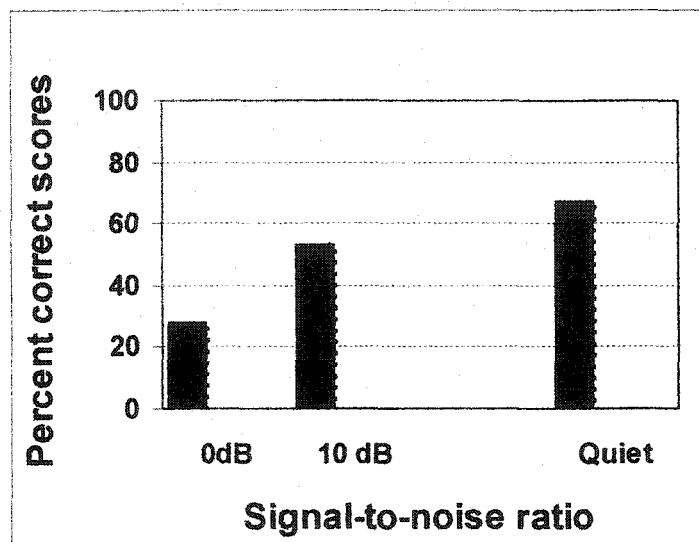
Figure 16. Mean consonant recognition performance of hearing-impaired adults as a function of the number of sine waves



Post-hoc testing ($<.01$ level of significance) reveals that the mean score for the unprocessed stimuli (58.3%) was significantly better than the scores for all other stimuli. The mean score for 4 sine waves (38%) was significantly poorer than all other conditions. The mean scores for 8 and 12 and 16 sine waves (48.5%, 49.7%, and 52.4% respectively) did not differ significantly.

Mean performance as a function of signal-to-noise ratio is shown in Figure 17. Inspection of the figure reveals that performance decreased as signal-to-noise ratio decreased. Post hoc analysis ($<.01$ level of significance) reveals significant differences between all three means obtained in quiet (67%), and +10 dB SNR (53%) and at 0 dB SNR (28%).

Figure 17. Mean Performance of hearing-impaired listeners as a function of signal-to-noise ratio.



The mean consonant recognition performance of hearing-impaired participants as a function of vowel environment is shown in Figure 18. Inspection of the figure reveals that the performance in the /a/ vowel environment was better than in the /i/ and /u/ environment. Post hoc analysis ($<.01$ level of significance) revealed that the mean recognition of consonants in the /a/context (54.7) was significantly better than mean performance in the /i/ context and /u/ context (46.1%, 47.4% respectively).

Figure 18. Mean consonant recognition performance of hearing-impaired listeners as a function of vowel context.

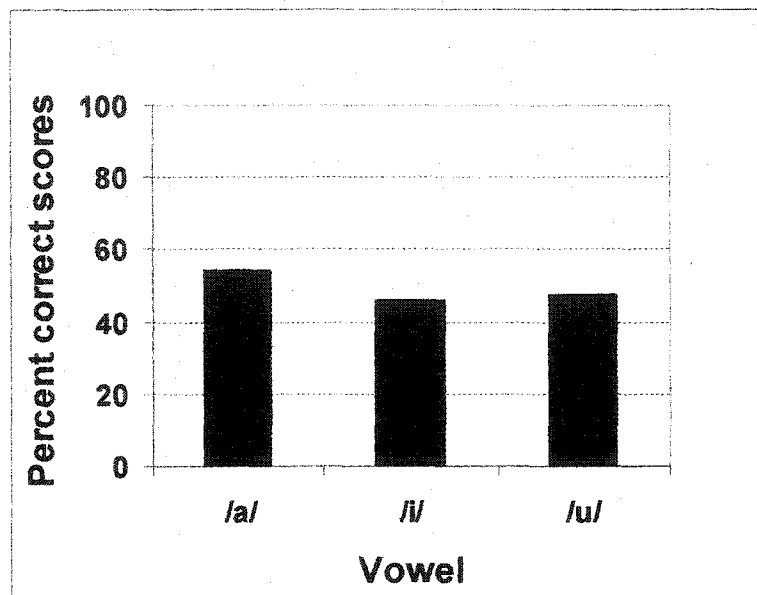


Figure 19 shows the mean performance as a function of the number of sine waves and signal-to-noise ratio. Inspection of the figure reveals that the mean recognition with 4 sine waves was significantly poorer than all other sine wave conditions at each signal-to-noise ratio. Recognition of the unprocessed stimuli was significantly better than that of the sine wave condition at each signal-to-noise ratio. Scores on the unprocessed stimuli also differed significantly as a function of signal-to-noise ratio. While performance asymptotes with 8 sine waves in the quiet and 0 dB SNR condition, a slightly different pattern was evident at +10 dB SNR. Post hoc analysis (<.01 level of significance) revealed that in quiet and at 0 dB SNR the means for 8, 12, and 16 sine waves were not significantly different. For +10 dB SNR the means with 8 and 12 sine waves, and the means for 12 and 16 sine waves are not significantly different.

Figure 19. Consonant recognition scores of the hearing-impaired group as a function of the number of sine waves and the signal-to-noise ratio.

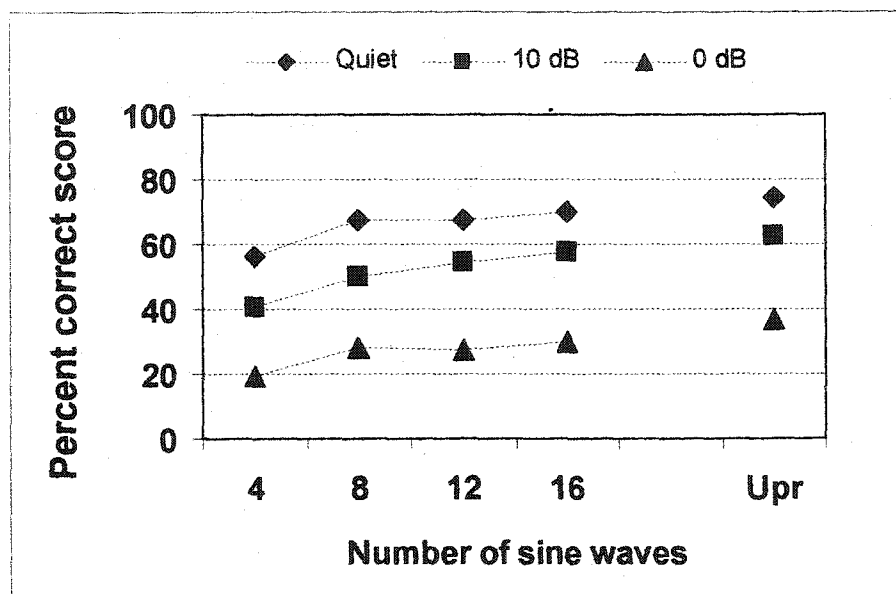
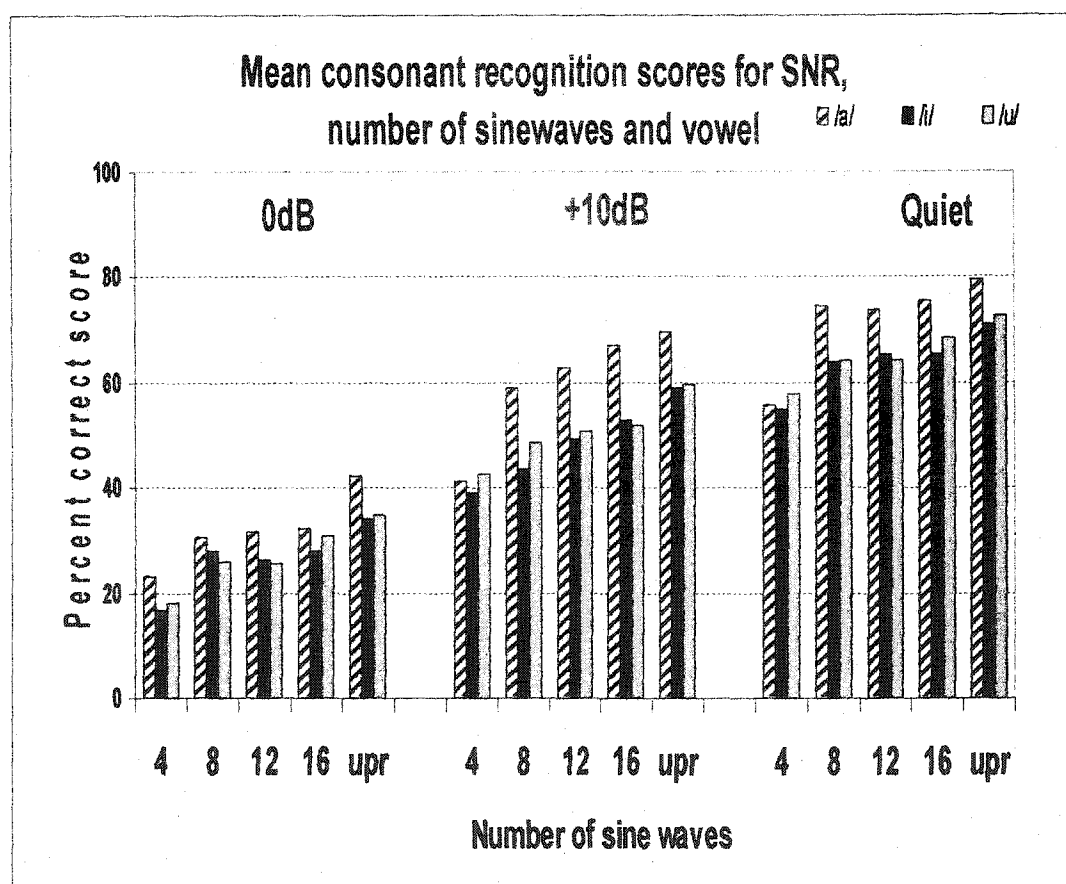


Figure 20 shows the significant interaction between the number of sine waves, signal-to-noise ratio, and vowel context. Overall the main findings are: In quiet there was no significant difference between recognition of stimuli with 8, 12, and 16 sine waves. There was a small but significant increase in performance with unprocessed stimuli. The mean recognition of consonants in the /a/ context was better than in the /i/ and /u/ contexts.

In the +10 dB SNR condition, no asymptote in performance was observed. Consonant recognition performance increased with increasing numbers of sine waves. The highest scores were achieved with the unprocessed stimuli. Recognition of consonants was significantly better in the /a/ context than in the /i/ and /u/ context for stimuli with 8 or more sine waves.

In the 0 dB SNR condition, overall performance was significantly lower than in the other signal-to-noise ratio conditions. Performance with /a/ was still better than with /i/ and /u/. The scores did not improve significantly with increasing numbers of sine waves. The performance with unprocessed stimuli was still significantly better than with sine wave stimuli for the vowel /a/ but no significant difference in performance was observed between 8, 12, 16, and the unprocessed stimuli for the vowels /i/ and /u/.

Figure 20. The mean consonant recognition performance for hearing-impaired adults in quiet as a function of signal-to-noise ratio, vowel, and number of sine waves



To summarize, for the hearing-impaired adults speech recognition of stimuli with 4 sine waves was significantly poorer than with 8, 12, or 16 sine waves. There was no significant difference in recognition among stimuli created using 8, 12, or 16 sine waves. Recognition of the unprocessed stimuli was consistently better than all of the sine wave stimuli. The vowel environment had a significant effect on consonant recognition. Recognition was significantly better in the /a/ context than with /i/ or /u/. The best performance was obtained in quiet. Performance decreased significantly as the signal-to-noise ratio was decreased to +10dB SNR, and was further decreased at 0 dB SNR.

4.3. Observations from the confusion matrices

Separate confusion matrices were generated for each sine wave and vowel condition for the normal hearing and for the hearing-impaired adults. The main findings are summarized below. The confusion matrices for normal hearing adults are shown as tables 1-15 in Appendix 2. The error patterns for hearing-impaired adults are shown as tables 1-15 in Appendix 3. The confusion matrices are shown for each sine wave and signal-to-noise ratio collapsed across vowels.

Overall both normal hearing and hearing-impaired adults showed errors of place and manner as the number of sine waves decreased. The most common substitution errors were plosives replaced by fricatives in the same voicing category. Some voiced plosives were substituted by the semivowels. Voiceless fricatives and affricates were substituted for voiceless plosives. Plosives and affricates in the same voicing category were substitute for fricatives. Affricates were resistant to errors, but when errors were made, plosives in the same voicing category were a common

substitution. Liquids and semivowels were substituted for nasals. Liquids were substituted for semivowels. Nasals and semivowels were substituted for liquids.

4.3.1. Quiet

For the unprocessed condition in quiet, normal hearing adults showed very few errors. Hearing-impaired adults made more errors. These were mainly errors of place and manner in the same voicing category.

With 12 sine waves normal hearing adults still showed very few errors, while hearing-impaired adults showed a greater number of errors. The error patterns for the hearing-impaired adults remained similar to those seen in quiet. As the number of sine waves decreased to four, errors for both the normal hearing and hearing-impaired groups increased. Normal hearing adults showed a large number of errors of place and manner but they were made within the same voicing category as the target sound. For hearing-impaired adults, the error patterns showed substitutions of manner and place extending across voicing categories. Overall hearing impaired adults made more errors than normal hearing adults. The vowel environment had little effect on consonant recognition in quiet for either group.

4.3.2. +10 dB signal-to-noise ratio

For normal hearing adults, performance was fairly good. Errors in recognition of unprocessed, 16, 12, and 8 sine waves were primarily place errors. When the number of sine waves was reduced to four, the number of errors increased rapidly. Errors included substitutions of manner and place and a few voicing errors were also observed. There was no vowel effect observed for the errors.

For hearing-impaired adults, there were far more errors recognizing the unprocessed stimuli in the quiet condition. Many of the errors were in recognizing

fricatives and sonorants. Place and manner substitutions were common. With 12 sine waves, incorrect responses were observed with regard to correct identification of voicing, especially for the voiced plosives and fricatives. More errors of manner were present across all categories. With four sine waves the errors were less predictable. There were substitutions of all categories and the overall number of errors increased.

4.3.3. 0 dB signal-to-noise ratio

At 0 dB SNR, speech recognition of normal hearing adults was poorer. The errors in identification included substitutions of manner, place and voicing. A vowel effect was observed, with more errors seen with /u/ than with /a/ and /i/. When the number of sine waves decreased to 12 sine waves, voicing errors were common, and further down to 4 sine waves the errors were random and included all possible manner, place and voicing categories. More errors were found in the /u/ context than in /a/ or /i/ contexts.

For the hearing-impaired adults as with the normal hearing adults, there were more errors at 0 dB SNR than at the other signal-to noise ratios. For the unprocessed stimuli, there were very few voicing errors in addition to place and manner errors. The errors increased with decreasing numbers of sine waves. The vowel effect showed /u/ as having more errors than /a/ and /i/. There was no specific pattern of errors at this point as the categories of manner, place, and voicing were all present in the responses across all the consonants. Overall the performance of hearing-impaired adults was poorer than the normal hearing adults.

4.4. Information Transfer Analysis

From the confusion matrices, the amount of information transmitted for the consonants grouped into voicing, manner and place were calculated. The information transfer statistic is a measure of covariance between input and output. It can be described as the transmission from a variable 'x' to a variable 'y' in bits per stimulus. The ratio varies from 0 to 1. If the response is not closely related to the stimulus then the transmission is said to be poor, and values will be closer to 0. If the response can be predicted from the stimulus, then the transmission index will be closer to 1 (Miller

and Nicely, 1955). A percentage value was used here for the transmission measures.

Separate information transfer scores were computed for the normal hearing and hearing impaired subjects.

Table 7.

Abbreviations for feature matrix in the analyses.

Voicing	Place	Manner
V = Voiced	Bil = Bilabial	Pls = plosive
Vs = Voiceless	Lad = Labiodental	Frc = Fricative
	Alv = Alveolar	Aff = Affricate
	Pal = Palatal	Nsl = Nasal
	Vel = velar	Lqd = Liquid
		Gld = Glide

Table 8.

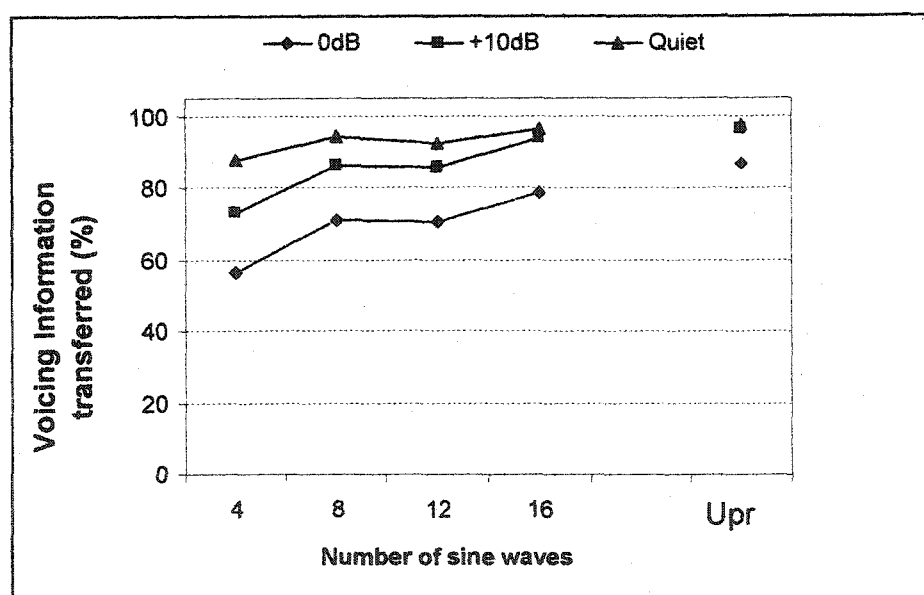
Matrix of consonants and abbreviations used to denote voicing, manner and place.

Consonant	p	t	k	b	d	g	f	θ	s	ʃ
Voicing	Vs	Vs	Vs	V	V	V	Vs	Vs	Vs	Vs
Place	Bil	Alv	Vel	Bil	alv	Vel	Lad	Alv	Alv	Pal
Manner	Pls	Pls	Pls	Pls	pls	Pls	Frc	Frc	Frc	Frc
Consonant	v	z	m	n	l	r	tʃ	dʒ	j	w
Voicing	V	V	V	V	V	V	Vs	V	V	V
Place	Lad	Alv	Bil	Alv	Alv	Alv	Alv	Alv	Pal	Bil
Manner	Frc	Frc	Nsl	Nsl	Lqd	Lqd	Aff	Aff	Gld	Gld

4.4.1. Information transfer for normal hearing adults

The percentage of voicing information transmitted for normal- hearing participants is shown in Figure 21. Examination of the figure reveals that the voicing information transmitted decreased with decreasing signal-to-noise ratio. In quiet, asymptotic performance was reached with 8 sine waves. At +10 dB SNR and 0 dB SNR the voicing information transmitted continue to grow with increasing number of sine waves.

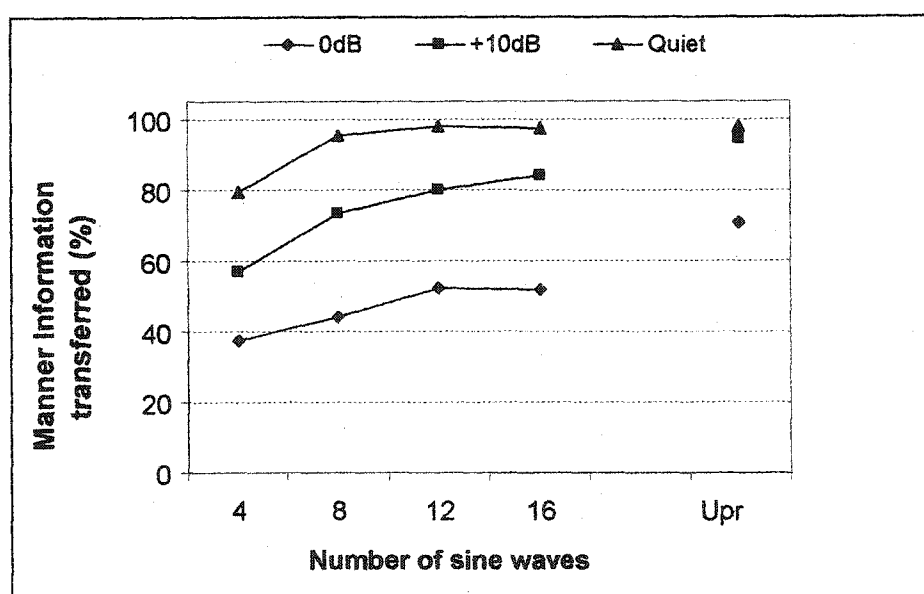
Figure 21. Voicing information transferred to normal hearing adults as a function of the number of sine waves and the signal-to-noise ratio.



The percentage of manner information transferred for the normal hearing adults as a function of number of sine waves and signal-to-noise ratio is shown in Figure 22. From the figure it can be observed that asymptotic performance was reached with 8 or more sine waves in quiet. The total manner information transmitted

by unprocessed speech in quiet was similar in quiet to transmission at +10 dB SNR. For +10 dB SNR and 0 dB SNR there was an increase in the manner information transmitted as the number of sine waves was increased. At all signal-to-noise ratios the manner information percentage transmitted was highest for the unprocessed condition.

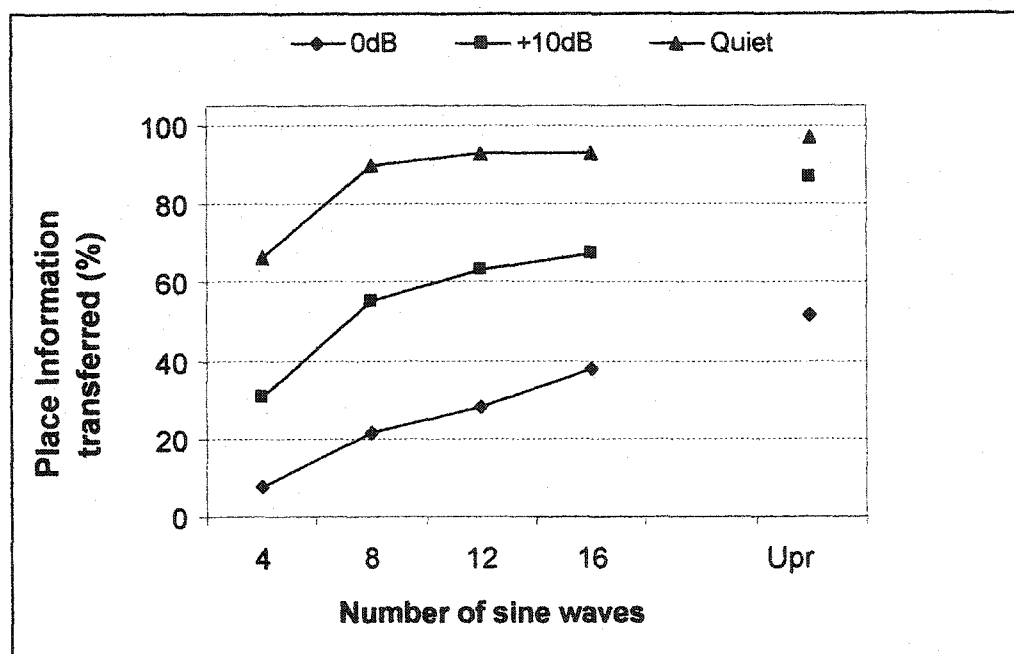
Figure 22. Manner Information transferred to normal hearing adults as a function of the number of sine waves and signal-to-noise ratio.



The percentage information transferred for place as a function of the number of sine waves and signal-to-noise ratio is shown in Figure 23. Examination of the figure reveals an increase in the transference of place information with increasing number of sine waves and increasing signal-to-noise ratio. In quiet, asymptotic transfer of place information was seen with 8 sine waves. For +10 dB SNR and 0 dB SNR the place information transmitted increased with increasing numbers of sine waves (and never reached the same level as for the unprocessed stimuli).

Unprocessed stimuli delivered the maximum amount of place information at all signal-to-noise ratios.

Figure 23. Place information transferred to normal hearing adults as a function of the number of sine waves and signal to noise ratio.

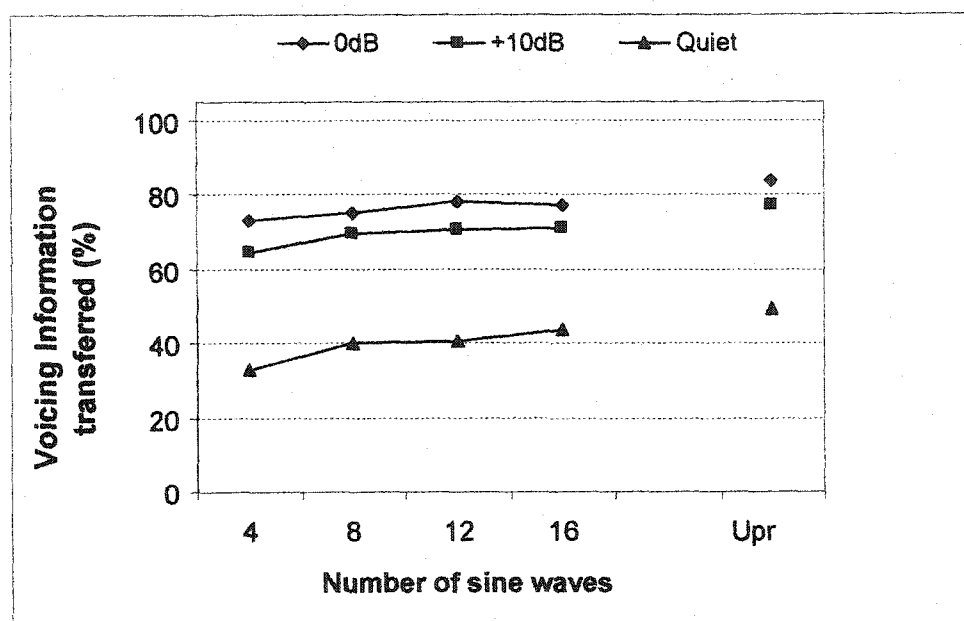


4.4.2. Information transfer for hearing-impaired adults

The percentage information transferred by voicing as a function of the number of sine waves and signal-to-noise ratio is shown in Figure 24. From the figure it can be observed that the percentage of information transferred decreased as signal-to-noise ratio decreased. There was a very small difference in information transfer between quiet and +10 dB SNR. The information transfer is substantially less at 0 dB SNR. The percentage of information transferred increased slowly with an increasing number of sine waves, although an increase from 12 to 16 sine waves did not appear

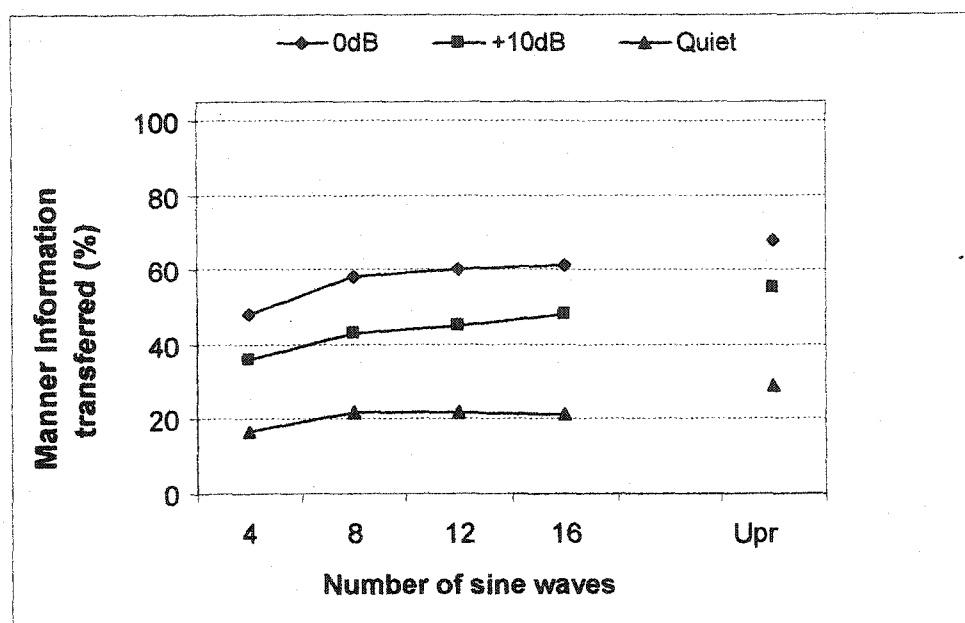
to be significant. Unprocessed stimuli showed a greater amount of voicing information transferred than the processed stimuli at all signal-to-noise ratios.

Figure 24. Voicing information transferred to hearing-impaired adults as a function of the number of sine waves and signal-to-noise ratio.



The manner information transferred for hearing-impaired participants as a function of the number of sine waves and signal-to-noise ratio is shown in Figure 25. The figure shows very little increase in percent manner information transferred for 8 or more sine waves. The unprocessed stimuli showed the highest percentage of information transferred at all signal-to-noise ratios. The information transferred decreased with decreasing signal-to-noise ratio.

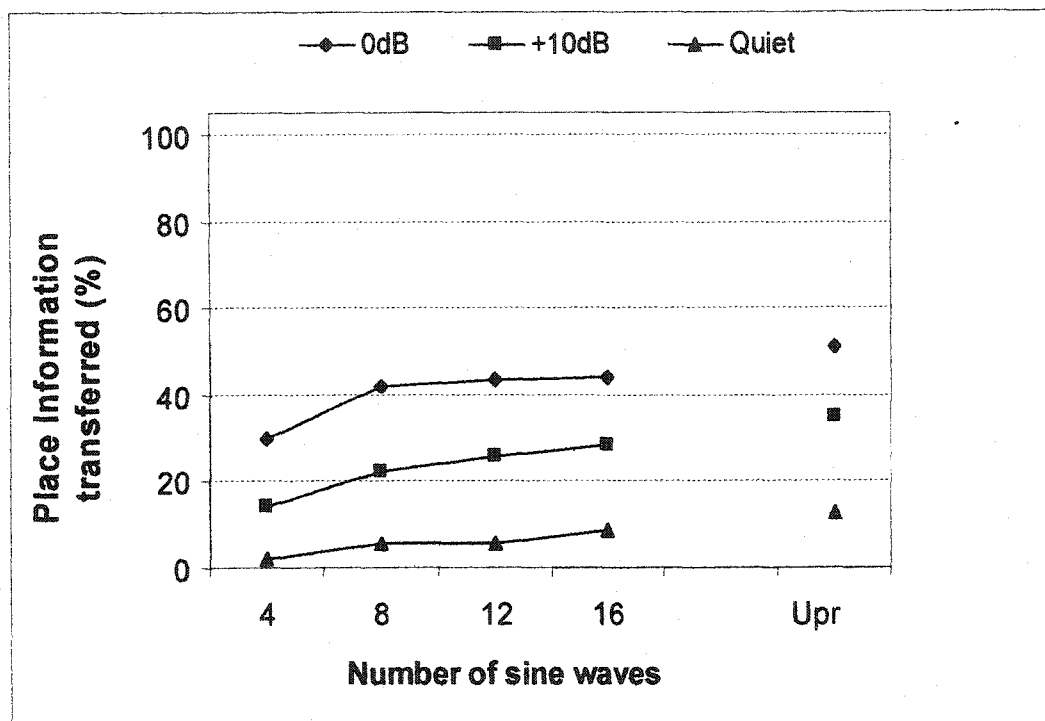
Figure 25. Manner information transferred to hearing-impaired participants as a function of the number of sine waves and signal-to-noise ratio.



The percentage of information transferred by place is shown in Figure 26.

From the figure it can be observed that information transferred by place shows small increments with increasing numbers of sine waves. Place information transferred decreased with decreasing signal-to-noise ratio. The maximum place information was transferred by the unprocessed stimuli at each signal-to noise-ratio.

Figure 26. Place information transferred to hearing-impaired participants as a function of the number of sine waves and signal-to-noise ratio.



In summary, normal hearing participants showed a much greater percentage of information transferred than did hearing-impaired participants. The difference in percentage of information transferred increased with decreasing SNR. The pattern of information transfer for voicing, manner and place was essentially the same for both the hearing-impaired and the normal hearing adults with one exception. The percentage information transferred was greatest for voicing for all conditions except for normal hearing adults in quiet. For this one condition, the percentage of information transferred was slightly less than that for manner cues.

With this one exception, voicing information was transferred very effectively even with very few sine waves, while manner information was transferred less effectively and place information was transferred least effectively.

Background noise also served to reduce the amount of information transferred. Maximum information was transferred for the quiet condition, and information transferred was progressively reduced with decreasing signal-to-noise ratio.

4.5. Results of the Acoustic Analyses

A wideband spectrogram (256 point FFT, sampling rate of 22050, and a Blackman-Harris window) was generated for all of the stimuli using a commercial wave editing software package (MultiSpeech). The carrier phrase was not included in the acoustic analyses.

Important acoustic features of consonant identity i.e., burst frequency, formant transitions, frication noise and voicing bar were identified for each consonant. Stimuli were examined to identify which acoustic features were preserved with decreasing numbers of sine waves. The effect of background noise on the acoustic features was also examined. Figure 27 shows the wideband spectrogram of the unprocessed syllable /aba/. Figures 28, 29, and 30 show the wideband spectrograms of the syllable /aba/ processed with 16, 8, and 4 sine waves respectively. As can be seen from the figures, as the numbers of sine waves decrease the high frequency energy in the stimuli progressively decreases. As shown in Figure 28, for /aba/ at 16 sine waves the burst of the plosive and the formant transitions are still clear. Figure 29 showing /aba/ at 8 sine waves shows very little burst information and marginal formant transitions. Figure 30 showing /aba/ with 4 sine waves shows no burst or formant transition information.

Figure 27. Wideband Spectrogram on the syllable /aba/ unprocessed in quiet
(Frequency scale is 0-8500 Hz).

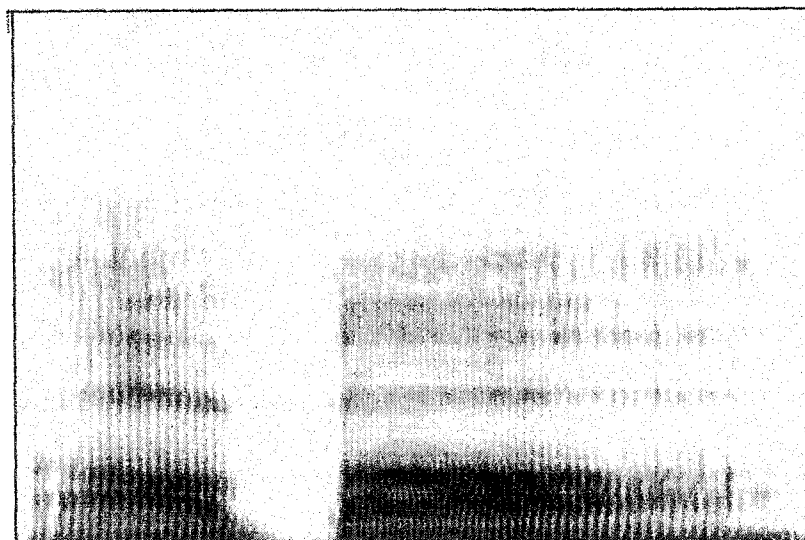


Figure 28. Wideband Spectrogram on the syllable /aba/ with 16 sine waves in quiet
(Frequency scale is 0-8500 Hz).

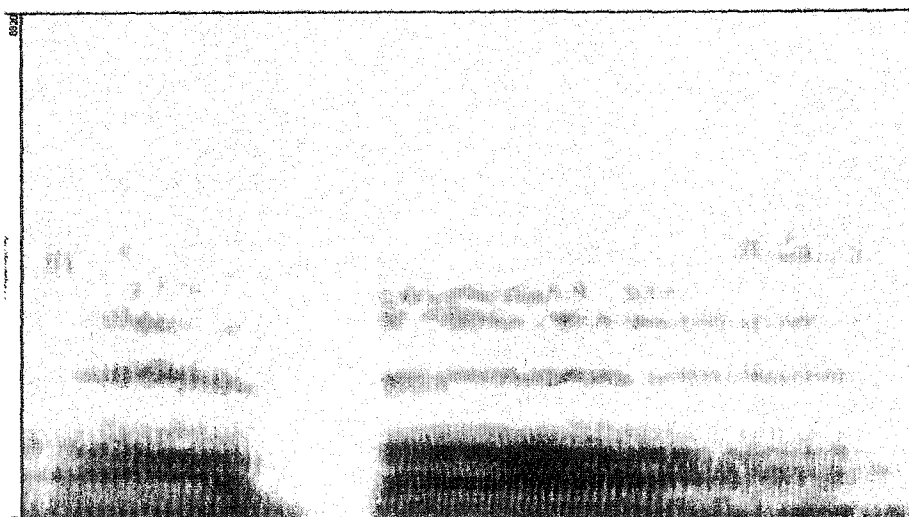


Figure 29. Wideband Spectrogram on the syllable /aba/ with 8 sine waves in quiet
(Frequency scale is 0-8500 Hz).

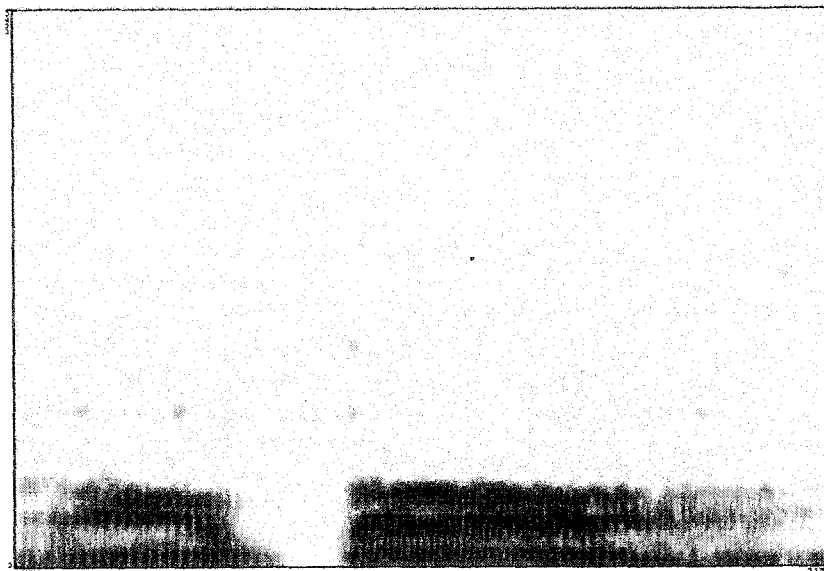


Figure 30. Wideband Spectrogram on the syllable /aba/ with 4 sine waves in quiet
(Frequency scale is 0-8500 Hz).



Figure 31 shows the syllable /asa/ modeled with 16 sine waves in quiet. The frication noise and formant transitions are clearly visible although a lot of high frequency information has already been removed.

Figure 32 and Figure 33 show the syllable /asa/ modeled with 16 sine waves in the presence of background noise at +10 dB SNR and 0 dB SNR respectively. In noise (+10 dB SNR) the features are more difficult to identify due to the background noise. Few of the acoustic features are observed at 0 dB SNR. From the figures it can be seen that noise tends to fill in the gaps between the speech information. Spectral definition is much poorer at 0 dB SNR than at +10 dB SNR. Appendix 1 shows the wide band spectrograms for the syllables /apa/, aba/, /asa/, /aza/, /atʃa/, /ana/, /ara/, and /aja/ at all sine wave and signal-to-noise ratio conditions.

Figure 31. Wideband Spectrogram on the syllable /asa/ with 16 sine waves in quiet (Frequency scale is 0-8500 Hz).



Figure 32. Wideband Spectrogram on the syllable /asa/ with 16 sine waves in +10 dB
SNR (Frequency scale is 0-8500 Hz).

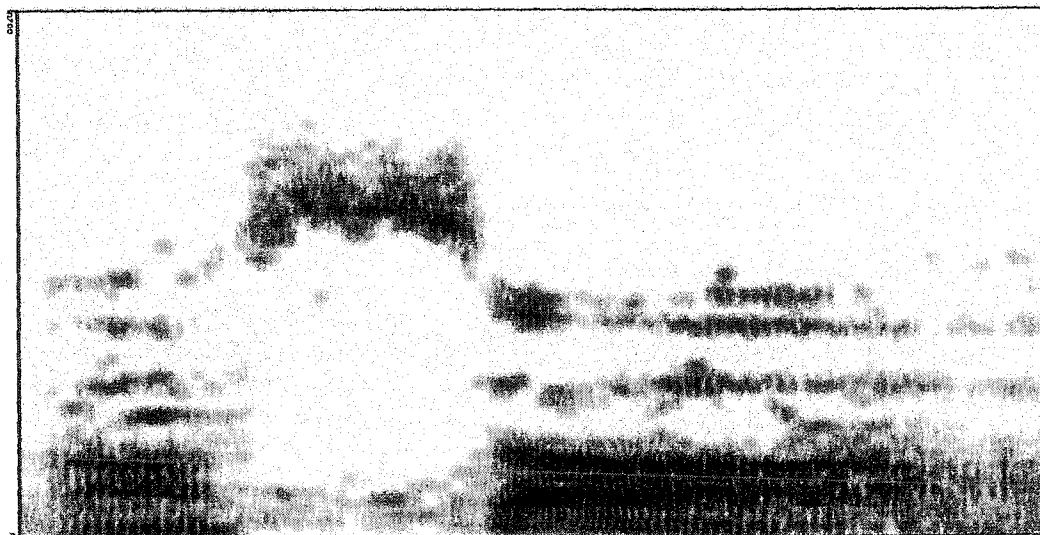


Figure 32. Wideband Spectrogram on the syllable /asa/ with 16 sine waves in 0 dB
SNR (Frequency scale is 0-8500 Hz).



A summary of the salient acoustic features preserved in the stimuli with different numbers of sine waves in the quiet condition are described below. Features are described for the unprocessed condition, for the 12 and 16 sine wave conditions grouped together, and for the 4 and 8 sine wave conditions grouped together. Any patterns that were characteristic for a particular vowel context are described under vowel effects.

Salient acoustic characteristics preserved in the stimuli with different numbers of sine waves in the quiet condition:

Consonant /p/: Bilabial Voiceless stop

Unprocessed:

Aspirations on all bursts

Burst and formant transition

12-16 sine waves:

Burst

Rising formant transitions clear

4-8 sine waves:

Burst

Transitions unclear

Vowel effects:

Formant information identifiable for 4 sine waves for /ubu/

Consonant /t/: Alveolar Voiceless stop

Unprocessed:

Well defined high frequency burst, and formant transitions

/t/: 12-16 sine waves:

Burst present and transitions clear

4-8 sine waves:

Some burst information,

Some F2 information

Formant transitions weak

Vowel effects:

Falling F3 transition clear even at 4 sine waves with /u/

Consonant /k/: Velar Voiceless stopUnprocessed:

Very diffuse burst, formant transitions visible

12-16 sine waves:

Burst energy; Strong transitions

4-8 sine waves:

Very little burst energy; weak but clear transition energy

Consonant /b/: Bilabial Voiced stopUnprocessed:

Voicing bar not clear

Very faint burst of a short duration; Rising F2 transitions

Differences between voiced and voiceless clearer based on gap
duration

12-16 sine waves:

Little burst energy; Transitions visible

/b/: 4-8 sine waves:

No burst energy; falling transitions from vowel to consonant

Very little information

Difficult to differentiate between /b/, /d/, and /g/ at 4 sine waves

Vowel effects:

Stronger burst with more diffuse energy for /u/ with F1 and F2 clear till 8 sine waves

Consonant /d/: Alveolar Voiced stopUnprocessed:

Uniform burst with falling formant transitions.

12-16 sine waves:

F2 clear; little burst energy

4-8 sine waves:

Burst information present for /i/ and /u/;

Transitions clear till 8 but not at 4

Vowel effects:

Burst and transition of F2 clear till 4 for /udu/

Consonant /g/: Velar Voiced stopUnprocessed:

Mid frequency burst with steep F2 transitions

12-16 sine waves:

Burst clear with falling F2 transitions

/g/: 4-8 sine waves:

Burst clear till 8 but not at 4; Transitions not clear

Vowel effects:

Clear spacing between F1 and F2 seen with /igi/ down to 4 sine waves

Consonant /f/: labiodental voiceless fricativeUnprocessed:

Very flat diffuse spectra with formant transitions

Very little difference from /θ/ at low numbers of sine waves

12-16 sine waves:

Very little frication energy and transitions

4-8 sine waves:

No frication energy and transitions; Only vowel formants

Consonant /θ/: dental voiceless fricativeUnprocessed:

Mid to high frequency frication energy with formant transitions

12-16 sine waves:

Frication energy present, transitions present

4-8 sine waves:

Frication energy present till 8; Transition energy present till 4

Consonant /s/: alveolar voiceless fricative

Unprocessed:

Well defined high frequency frication noise; Transitions clear

12-16 sine waves:

Clear frication energy and formant transitions

4-8 sine waves:

Frication energy identifiable

Transitions of F1 and F2 are seen

Consonant /ʃ/: palatal voiceless fricative

Unprocessed:

Robust high frequency energy and transitions

12-16 sine waves:

Well defined frication energy and formant transitions

4-8 sine waves:

Transition and frication energy clear till 4 sine waves

Vowel effects:

Flatter formant transitions with /i ʃ i/

Consonant /v/: labiodental voiced fricative

Unprocessed:

Little frication energy; clear formant transitions

12-16 sine waves:

Transitions clear

/v/: 4-8 sine waves:

Formant transitions present at 4. No frication energy present

Vowel effects:

High frequency spectral energy in unprocessed for /uvu/

Consonant /z/: alveolar voiced fricativeUnprocessed:

More diffuse energy than /s/. Transition clear

12-16 sine waves:

Frication energy and formant transitions present

4-8 sine waves:

Little fricative and transition energy at 4, though clear at 8

Voicing bar is present with 16, 12, and 8 sine waves

Vowel effects:

Transition energy weak at 8 for /uzu/

Consonant /tʃ/: alveolar voiceless affricateUnprocessed:

Broad spectrum, mid to high frequency noise with clear transitions

12-16 sine waves:

Vowel to consonant rising F2, and consonant to vowel falling F2

Frication noise

4-8 sine waves:

Clear noise energy and formants down to 4

Consonant /dʒ/: alveolar voiced affricate

Unprocessed:

Voicing bar unclear

Gap duration between voice and voiceless affricate clear

Broad noise band with clear formant transitions

12-16 sine waves:

Clear noise energy and falling formant transition of F2

4-8 sine waves:

Little noise energy and poor transitions, more plosive like

Consonants /m/: Bilabial nasal and /n/: alveolar nasal

Unprocessed:

Clear anti-resonances and formant transitions

12-16 sine waves:

F1 and F2 are present. High frequency information decreases

4-8 sine waves:

F1 and F2 faint at 8 but not identifiable at 4

Spectrum resembles that of plosives

Consonants /r/ and /l/: Liquids

Unprocessed:

Clear F3 transition, falling for /r/

12-16 sine waves:

Clear transitions

/l/ and /r/: 4-8 sine waves:

Good transitions till 8 for /i/ and /u/

No F3 at 4

Vowel effects:

Difficult to tell /l/ from /r/ at 4 sine waves for /i/ and /u/

Consonants /w/: Bilabial semivowelUnprocessed:

Clear formant transitions of F1 and F2

12-16 sine waves:

Clear transitions

4-8 sine waves:

Formants are clear while transitions are unclear

Vowel effects:

Very little difference between F1 and F2 for /u/

Consonants /y/: Palatal semivowelUnprocessed:

Good formant structure; Low F1

12-16 sine waves:

Falling F2 rising F3 clear

4-8 sine waves:

Clear transitions

Vowel effects:

Unclear patterns with /u/ at 4

CHAPTER 5

Discussion

The main purpose of this study was to determine how varying the number of sine waves in sinusoidally modeled speech would affect consonant recognition in quiet and in the presence of background noise for normal hearing and hearing-impaired listeners. Speech was modeled using 4, 8, 12, and 16 sine waves. It was hypothesized that sine wave modeling would improve speech intelligibility in the presence of background noise for hearing-impaired individuals. Differences in error patterns between unprocessed speech and sine wave modeled speech at different signal-to-noise ratios, for both normal hearing individuals and hearing-impaired individuals were also determined.

The main findings of the study were:

- 1) In quiet, consonant recognition did not improve significantly by increasing the number of sine waves above 8.
- 2) In noise, consonant recognition continued to increase as the number of sine waves were increased up to 16 sine waves.
- 3) Decreasing the signal-to-noise ratio reduced the intelligibility of sine wave modeled consonants.
- 4) Vowel environment had a small effect on the intelligibility of sine wave modeled consonants.

- 5) As the number of sine waves was decreased, the transmission of speech feature information decreased, with the transmission of voicing being more robust than that of manner or place.
- 6) The performance of hearing-impaired listeners was poorer than that of normal hearing listeners under all listening conditions.
- 7) The unprocessed consonants were always more intelligible than the sine wave modeled consonants.

In the discussion that follows, the effect of the number of sine waves, signal-to-noise ratio and vowel environment are discussed.

5.1. Number of Sine waves

In this study, speech was modeled using 4, 8, 12, and 16 sine waves. Consonant recognition of the sinusoidally modeled speech was compared to recognition of the unprocessed stimuli.

A similar pattern of results was obtained for both normal-hearing and hearing-impaired adults. Overall, the poorest consonant recognition scores were obtained with 4 sine waves, and the best consonant recognition scores were obtained with unprocessed stimuli. For the stimuli processed in quiet, asymptotic performance was reached with 8 sine waves. The scores for the unprocessed stimuli were always higher than the scores for all of the sine wave conditions.

Normal hearing adults showed a mean consonant recognition score of 86% with 16 sine waves and a mean score of 93 % for the unprocessed stimuli. Hearing-impaired adults showed a mean consonant recognition score of 52% with 16 sine waves and 58% for the unprocessed stimuli. While the overall scores for the hearing-impaired group were lower than those of the normal hearing group, both groups showed similar differences between the best sine wave condition (16 sine waves) and the unprocessed signal (approximately 6-7 percentage points).

The analysis of transmission of features for normal hearing subjects showed that in quiet, features of voicing, manner and place almost reached asymptote with 8 sine waves. Very small increments were seen with additional sine waves. Once again, there was a very small difference with the features transmitted for the unprocessed signal. The same was true for the hearing-impaired adults, although the overall percentage of features transmitted was lower than that for the normal hearing adults. Thus, in quiet, it appeared that most of the important features for consonant recognition were present in the information carried by the first 8 sine waves.

5.1.1. Acoustical Analysis from spectrograms

The spectrographic analyses supported the consonant recognition data, the transmission features analysis, and the consonant confusion data which indicated that important features for consonant recognition across all classes of sounds were coded within the first eight sine waves. More information was added at 12 and 16 sine waves, but information critical to identifying the consonant was present at 8 sine waves. Consonant confusions revealed that for normal hearing adults, very few errors were

observed in the quiet condition. When speech was modeled using 16 and 12 sine waves the errors were those of place. When speech was modeled using 8 and 4 sine waves manner errors emerged as well. For speech modeled using 4 sine waves, occasional voicing errors were seen. Hearing-impaired individuals showed similar patterns of errors in quiet to those of the normal hearing adults, but also showed a larger number of errors.

The spectrographic analysis revealed that as the number of sine waves decreased from 16 to 4, the high frequency energy in the signal decreased, thereby losing cues to place of articulation. The temporal structure remained essentially the same as the numbers of sine waves were reduced from 16 to 8 and, as a consequence, manner cues were affected to a much lesser extent than place cues. With 8 and 4 sine waves, there was little burst information. There was also marginal formant transition information for the two conditions (4 and 8 sine waves), although the formant structure of the vowels was still well preserved (except for the 4 sine wave condition where only two formants were evident).

Much of the following discussion will focus on the 4 and 8 sine wave conditions. For stop consonants, much of the noise burst was missing as well as the second formant transition so that manner cues were reduced substantially with 4 sine waves. Voicing cues remained relatively unaffected as the number of sine waves was reduced.

The affricates showed a strong spread of mid to high frequency energy and robust formant transitions which were well maintained down to 4 sine waves. They were identified with the greatest accuracy of all the consonant classes. The voiced/voiceless distinction was also distorted for the 4 sine wave condition in much the same way as for

the stop consonants. This might explain the voicing errors at low numbers of sine waves for this group.

The sibilants (/s/, /z/, /ʃ/) showed a gradual decrease in frication energy with decreasing numbers of sine waves. The voiceless sibilants showed frication energy down to 4 sine waves, while voiced sibilants showed a combination of frication energy and formant transitions right down to 4 sine waves.

The non sibilants (/f/, /θ/, /v/) showed formant transitions down to 12 sine waves, but lost definition at 8 sine waves. At 4 sine waves, the spectrograms of the different consonants resemble each other within the same vowel set. This explains the large numbers of confusions within the fricative category at the 4 and 8 sine wave conditions.

Nasals, liquids and glides were found to be easily confusable with decreasing numbers of sine waves. The process of sine wave modeling systematically removes portions of the amplitude spectrum with decreasing numbers of sine waves starting with the least intense components. The formant transitions, especially those of F3, play an important role in the identity of these consonants. The transitions are softer in intensity than the lower formants and therefore get removed early in the process of sine wave modeling as the number of sine waves decrease. Hence for the group of speech sounds consisting of the nasals, liquids, and glides, formant transitions are not robust as the numbers of sine waves decrease. As a result, at very low numbers of sine waves the spectrograms of this group of consonants look identical. This would explain why in the confusion matrices at low numbers of sine waves the nasals, liquids and glides are easily substituted for one another.

5.1.2. Temporal Cues

An additional set of cues that is not clearly evident from the spectrograms is that contained in the overall temporal envelope. Phonetic information conveyed by the temporal envelope includes manner cues, such as the presence of periodicity and aperiodicity. Fluctuations in the amplitude waveform also cue first and second formant identity and thereby place, and periodicity information cues voicing (Rosen 1992). Some of the envelope cues are still preserved in the waveform of the stimuli with 4 sine waves.

In summary, decreasing the number of sine waves eliminates the weaker components of the speech signal, effectively removing the formant transitions and high frequency cues that convey place information. While the loss of spectral detail does affect consonant intelligibility, temporal cues, such as burst duration and voice onset time are more robust and are affected less by a reduction in the number of sine waves except in the 4 sine wave condition. As a consequence, manner and voicing cues are relatively robust and are reduced only moderately by sine wave modeling, except for the extreme case of 4 sine waves.

5.1.3. Comparison to low-pass filtering and masking

The effect of sine wave modeling can be compared to that of low-pass filtering and masking. Decreasing the number of sine waves had similar effects to those obtained by Miller and Nicely (1955) by decreasing the cut-off frequency of a low pass filter or by adding noise. In both low pass filtering and in sine wave modeling, high frequency information is systematically removed from the speech spectrum. While removing high frequency components is intentional in low-pass filtering, it is incidental in sine wave

modeling. The lower intensity components in speech typically tend to be higher in frequency and get removed during the peak picking process.

Studies of the effects of low pass filtering (Miller and Nicely 1952, Wang, Reed, and Bilger, 1978) have found that place cues are affected first by low pass filtering followed by manner cues and voicing. Errors with low pass filtering were predictable depending on the high frequency energy being removed. Hence sibilants and plosives were affected first followed by affricates, and then nasals, glides and liquids. In the present study, as the numbers of sine waves decreased, place of articulation errors emerged first, followed by errors in nasality, affrication, duration (/s/, /j/, /z/) and finally voicing. Duration and affrication were very robust in both this study and that of Miller and Nicely. The error patterns observed with sine wave modeling in quiet with 12-16 sine waves were very similar to the error patterns observed by Miller and Nicely (1955) for stimuli that were low pass filtered at 250-2500Hz. They showed mainly place and few manner errors. The error patterns for the eight sine wave condition were comparable to the patterns seen for stimuli that were low-pass filtered at 250-1200 Hz in that more errors of place and manner were seen while voicing errors were infrequent. The error patterns for 4 sine waves were similar to those obtained for low pass filtered stimuli at 200-400Hz, in that errors of place, manner and voicing were common. Nasality was also significantly affected with 4 sine waves.

5.1.4. Effect of Signal Processing

The results of this study suggest that the information contained in the eight highest peaks of the speech signal is critical to consonant recognition. Speech modeled

with fewer than eight sine waves resulted in poorer consonant recognition. Adding more sine waves to the stimuli did not increase consonant recognition significantly. Although in the present study speech was not modeled with more than 16 sine waves, pilot studies did not show a difference in performance between 16, 20, and 24 sine waves. Thus, continuing to increase the number of sine waves in the signal may be redundant.

The unprocessed speech signal still shows the best transmission scores. It is likely that sinusoidal modeling introduces a type of distortion into the signal that limits speech recognition performance. In the analysis portion of the signal processing, multiple FFT's are generated, the peaks are identified and speech is resynthesized from the FFT information. This processing introduces some distortion in the signal which might affect speech intelligibility. The fact that there was a similar reduction in intelligibility for both the hearing-impaired and normal hearing group is consistent with an external source of distortion over a reduction of perceptual cues. This hypothesis would need to be tested by using speech that had undergone the analysis for sinusoidal modeling, but with all components preserved from the FFT analysis window.

5.2. Effect of signal-to noise ratio

In the current implementation of sine wave modeling, the process progressively removes low intensity components as the number of sine waves decreases. The initial hypothesis was that removal of the low-to-mid amplitude spectral information would also remove a large number of the noise components and thereby improve speech intelligibility. However, the results do not support that hypothesis.

The effect of noise on recognition of sine wave modeled speech was determined by comparing performance in quiet, +10 dB SNR, and 0 dB SNR. Consonant recognition

performance decreased with decreasing signal-to-noise ratios. The overall reduction was on the order of 11% from quiet to +10 dB SNR, and approximately 24% from +10 to 0 dB SNR for normal hearing adults. For hearing-impaired adults, the decrease in speech intelligibility was approximately 14% going from quiet to +10 dB SNR, and approximately 25% going from +10 to 0 dB SNR. There was no difference between the mean recognition of speech modeled with 8, 12, and 16 sine waves for normal hearing and hearing-impaired individuals at 0 dB SNR. At + 10 dB SNR, there was a slight improvement in performance seen with increasing numbers of sine waves for both groups.

The processed speech-in-noise did sound less noisy, but the reduction in noise did not translate into improved speech recognition. There are several explanations for why this did not occur. First, if the level of the background noise is high relative to speech, the temporal gaps in speech are filled in by the noise and they are picked by the processing algorithm as peaks. Second, the noise used in this experiment was cafeteria noise whose long term average spectrum matches the long term average speech spectrum. When cafeteria noise is mixed with speech and subjected to sine wave modeling, at any given instant the selected peak could belong to the noise or to speech. The reconstructed signal is thus composed of sine waves representing both speech and noise. When only 4 or 8 peaks are used in the sine wave modeling process, there may not be enough speech information represented by the sine waves to cue consonant identity. As more peaks are used for sine wave modeling, the chance that peaks from the speech will be picked is increased. Therefore speech intelligibility increases with an increasing number of sine waves in noise.

The second assumption was that the addition of noise masks the lower amplitude information in the speech signal and therefore the removal of this information (by the sine wave modeling process) would not decrease intelligibility. The results however indicate that the removal of these components actually decreases intelligibility. Speech recognition scores with the unprocessed speech-in-noise were superior to the sine wave modeled conditions. It is possible that information present at low amplitudes still contributes to consonant intelligibility and that not all of it is rendered unusable by background noise. This finding is in agreement with Turner and Henry (2003) who found that hearing-impaired listeners were able to utilize low amplitude high frequency information in the presence of background noise to improve speech intelligibility. Thus in quiet, a listener may not make use of the redundancies in the speech signal, but in the presence of background noise these cues become salient.

5.2.1. Analysis of Transmission of Features

In noise, the information transmitted continuously increased with increasing numbers of sine waves. The maximum level of information transmitted was achieved only with unprocessed stimuli. Information transfer functions for voicing, manner and place reveal similar patterns, but poorer levels of performance for decreasing signal-to-noise ratios. As the signal-to-noise ratio decreased, the largest decrements were observed in transmission of place information, followed by manner and finally voicing.

Overall, hearing-impaired individuals make more errors than normal hearing individuals both in quiet and in noise. At 0 dB SNR there were large reductions in place, manner and voicing information. At +10 dB SNR, voicing cues were still available, but there were reductions in manner and place information. As the number of sine waves

increased, performance in noise improved, but not to the same extent as for the normal hearing subjects. It appears that hearing-impaired individuals cannot use place and manner information in noise as effectively as normal hearing listeners.

5.2.2. Acoustic Analysis

Analysis of the spectrograms of the sinusoidal modeled stimuli reveals that addition of noise fills the temporal gaps in the speech. As a consequence, important information signaled by variation in signal level (e.g. the closure for stops, the anti-resonances for nasality) would no longer be salient cues. The presence of background noise also obscures and masks weak intensity cues such as formant transitions, burst energy for stops, and frication energy for non sibilants, therefore making identification of the consonants difficult. The consonant confusions support these findings. Place errors were made by the hearing-impaired listening to unprocessed stimuli at +10 dB SNR. With decreasing numbers of sine waves, manner and voicing errors were also made. Performance in the 0 dB SNR condition included voicing errors and errors across all manner and classes of consonants when the number of sine waves decreased.

5.2.3. Comparison to masking studies

In quiet, the effects of the loss of acoustic information due to the current implementation of sine wave modeling were very similar to the effects of adding background noise to the signal. Both sine wave modeling and masking tend to affect the perception of low intensity cues. The consonant recognition performance for the unprocessed stimuli in noise can be compared to studies of consonant recognition in the presence of background masking noise (Miller and Nicely, 1955, Dubno and Levitt, 1981). The deleterious effect of background noise on speech intelligibility is due to

several factors. First, noise masks the low intensity place cues in speech and therefore affects speech intelligibility. Second, low frequency components mask the lower intensity high frequency spectral information by the process of upward spread of masking, which affects both place and manner cues. Finally, noise fills in the temporal gaps, reducing the temporal information available to the listener. Background noise also masks low intensity components in the speech signal.

With the addition of background noise to the unprocessed stimuli, normal hearing adults showed errors of place in the +10 dB SNR condition with very few manner and no voicing errors. When the signal-to-noise ratio was decreased to 0 dB, a large number of errors were seen for place, manner and voicing. Similar patterns were observed by Miller and Nicely (1955), where at +12 dB SNR normal hearing listeners showed place errors and few manner errors, but showed errors of place, manner and voicing at 0 dB SNR. Dubno and Levitt (1981) found that at +5 dB signal-to-noise ratio, listeners made errors on consonant place, manner and voicing. Dubno and Levitt attribute this error pattern at poor signal-to-noise ratios to random guessing. The same pattern of errors (including errors involving voicing) was observed in this study when the number of sine waves decreased to 4, or at 0 dB SNR and when the overall scores were poor.

In the present study, transmission of information for nasals was affected in the +10 dB SNR condition, earlier than suggested by Miller and Nicely (1955). This finding is in agreement with the findings of Dubno and Levitt (1981) and in contrast with Miller and Nicely (1955) who found that transmission of nasals was relatively unaffected until -6dB SNR. Cafeteria noise was used in both the present study and the Dubno and Levitt study, rather than the random noise used by Miller and Nicely (1955).

The performance of hearing-impaired adults for the unprocessed condition in quiet was poorer than that of normal hearing subjects. Both normal hearing and hearing-impaired subjects showed similar reductions in percent correct scores due to the sine wave modeling process. With the addition of background noise, hearing-impaired adults showed larger decrements in performance from the unprocessed than did normal hearing adults. The errors made by hearing-impaired subjects were similar in pattern to the normal hearing subjects; however hearing-impaired individuals made a larger number of errors.

5.3. Effect of Vowel Environment

For both normal hearing and hearing-impaired listeners there was a significant difference in consonant recognition as a function of vowel environment. Performance in the /a/ context was better than in either the /i/ or /u/ context. This result was similar to that obtained by Dubno and Levitt (1981) for nonsense syllables presented in cafeteria noise. The vowel /a/ has its first two formants in the low to mid-frequency region and hence most formant transitions would have a gentle slope. This could be a reason for the more robust consonant recognition performance with /a/. The vowel /u/ has very low frequency, narrowly spaced first and second formants and this could give rise to steeper consonant transitions, and less separation between formants being retained by the sine wave modeling process. Recognition in the vowel /i/ context was better than /u/ for normal hearing subjects. This might be due to a larger separation of formants F1 and F2 and hence the presence of more identifiable information following resynthesis. Hearing-impaired adults may not have been able to utilize the high frequency second formant information due to poor frequency resolution and also due to upward spread of masking.

Their consonant recognition was similar in /i/ and /u/ environments, possibly because the first formant is the same for both vowels.

5.4. Comparison to Literature

5.4.1. Sine wave modeling

The findings of this study confirm and extend the findings of Kates (1994) that sinusoidal modeling can cause degradation of consonant recognition under certain listening conditions. Kates confined his testing to persons with normal hearing, whereas in the present study both persons with normal hearing and persons with hearing loss were tested. The hearing-impaired adults in this study showed similar patterns of performance to the normal hearing adults, although they did show more errors overall. Their error patterns suggest that the differences between normal hearing and hearing-impaired adults are more quantitative than qualitative. When degradation occurs, the consonant confusion analyses of both studies showed similar patterns. Information about voicing is more robust than information about manner and place of articulation.

Although the tests in noise were not conducted at similar signal-to-noise ratios (+5 dB for Kates), the results for the noise conditions (+10dB and 0dB) show similar decrements as those in Kates' study. The overall scores for the normal hearing subjects in this study were at least 10% higher than those in Kates' study. The choice of stimuli for this experiment could explain the discrepancy in scores. VCV syllables were used in this study while Kates used C+/a/, /a/+ C, and C + /i/. VCV syllables provide transitions into and out of the vowel, and can be said to have twice the number of cues to the identity of the consonant as compared to CV or VC syllables. Differences in the implementation of sinusoidal modeling might also account for some of the difference. The overlap-add procedure was used during the synthesis portion of the sine wave modeling applied in this

study, while Kates used a linear interpolation of successive time windows. The overlap-add procedure provides a perceptually superior signal with fewer distortions and this may also have contributed to the overall high performance on this study.

5.4.2. Comparison to cochlear implant processors

The findings of this study are similar to those that have emerged from research on processing strategies to be used in speech processors for cochlear implants.

Sine wave modeling encodes gross changes in the temporal envelope, as well as encoding high amplitude frequency information. The cochlear implant research has shown that processing strategies that preserve the changes in the temporal envelope are able to maintain speech intelligibility even in the absence of temporal fine structure (e.g., Dorman and Loizou, 1998; Loizou, Dorman and Tu, 1999; Dorman, Loizou, Fitzke and Tu, 1998; Fu, Shannon and Wang, 1998). Presenting the peaks in the short-term spectrum can preserve the components important for speech perception. The current findings are also in agreement with the cochlear implant research that suggests that in the presence of background noise more spectral information is needed to identify speech. For the cochlear implant processors, the number of channels needed to identify speech in the presence of background noise was increased from 6 in quiet to 8 in noise. In the present study, the consonant recognition scores continued to increase in the presence of background noise as the number of sine waves increased.

5.5 Comparison to spectral sharpening studies

The consonant recognition performance with unprocessed speech signal was always found to be better than with the sine wave modeled tokens. Bunnell (1990) and Simpson and colleagues (1990) had found that spectral sharpening yielded small improvements in speech intelligibility, while Stone and Moore (1992) did not find any improvements in intelligibility. It appears that very small improvements are seen for spectral sharpening which are not consistently observed. If any spectral sharpening was achieved by this implementation of sine wave modeling, the benefits were not observed as improvements in consonant recognition performance in this study.

5.6 Future directions for research

In the current implementation of sine wave modeling, peaks were picked in order of diminishing amplitude from the amplitude spectrum. The effects of this implementation were similar to the effects of low pass filtering on consonant recognition. The error patterns seen were a function of decreasing high frequency energy in the signal.

A different implementation of sine wave modeling with different methods for peak picking would be expected to provide different patterns of results. If the speech spectrum were whitened before peak-picking, more high frequency information would be present in the speech signal and would therefore give rise to a different pattern of recognition performance.

Speech recognition in the presence of background noise improved as the number of sine waves was increased. Under adverse listening conditions, both normal hearing and hearing-impaired listeners were able to utilize more high frequency energy to

improve speech recognition performance. It would be interesting to see if whitening the spectrum before picking peaks would provide further improvements in speech recognition performance in noise. Different strategies of peak picking, e. g., by frequency or by critical bands, would also give rise to different patterns of performance.

A limitation of an amplitude-based peak-picking algorithm is that for noise whose long-term spectrum is almost identical to that of speech, at poor signal-to-noise ratios, both speech and noise have an equal chance of being selected. In these cases an amplitude-based peak-picking strategy can only be effective at relatively good signal-to-noise ratios. For other types of noise the spectrum of the noise would need to be considered before using an amplitude-based peak-picking strategy.

More research needs to be carried out to explore which peak picking strategies will provide maximum speech intelligibility.

Consonant recognition performance was studied in the vowel contexts of /a/, /i/, and /u/. These three cardinal vowels were chosen as they represent three extremes positions in the vocal tract. The first and second formants of /a/ and /u/ are close in frequency, and the second and third formants of /i/ are close in frequency. It is possible that the patterns of vowel formants interacted with the peak picking algorithm to provide unique vowel effects. Further studies using a larger vowel set to include vowels with more widely separated formants are required to further define the effect of vowels on sine wave modeling.

CHAPTER 6

Summary and Conclusions

The main objective of this study was to determine how varying the number of sine waves in sinusoidally modeled speech would affect consonant recognition in quiet and in noise by normal-hearing and hearing-impaired adults in quiet and in the presence of background noise.

Sine wave modeling encodes the peaks of the original speech signal yielding an encoded signal that does not contain the lower amplitude components of the original speech signal. For speech in quiet, the speech can be modeled quite well with relatively few sine waves without causing significant degradation to speech recognition. For normal hearing adults listening in quiet, excellent (94%) consonant recognition was achieved with as few as 8 sine waves. For hearing-impaired adults listening in quiet, mean consonant recognition was close to maximum (67.5%) with 8 sine waves. The best score was obtained in the unprocessed condition (74.2%).

When noise was added to the signal before processing, the number of sine waves needed for consonant recognition increased with decreasing signal-to-noise ratio.

Although for normal hearing adults there was no difference in consonant recognition performance between the unprocessed (97%) and the 16 sine wave version (96%) when listening in quiet, in noise there were significant differences between the unprocessed condition and the modeled speech. At +10 dB SNR, mean recognition with 16 sine waves was 87%, but recognition of the unprocessed signal was 92%. At 0 dB SNR, a

mean recognition score of 61% was obtained with 16 sine waves, but the mean score for the unprocessed condition was 80%.

The same pattern of performance was evident for the hearing-impaired listeners, but the overall level of performance was poorer both for the unprocessed and the sine wave modeled speech. Mean recognition in noise at +10 dB SNR was 57% with 16 sine waves and 62% for the unprocessed condition. At 0 dB SNR, recognition with 16 sine waves was 30% and 37% in the unprocessed condition.

An information transfer analysis was helpful in explaining the pattern of errors made in quiet and in noise for sine wave modeled speech. As the number of sine waves used to encode the speech is reduced, information about the place of articulation is lost first, then manner, and finally voicing. This pattern of change in information transfer is similar to what occurs following low pass filtering. In quiet, 8 sine waves preserve most of the information about voicing, manner, and place of articulation.

A spectrographic analysis of the speech in quiet revealed that the acoustic cues to consonant identity that are present with 8 sine waves include formant transitions, some plosive burst energy, frication energy, and temporal cues such as gap duration. In quiet, both normal hearing, and hearing-impaired individuals are able to utilize this information in consonant identification. However, the overall performance of hearing-impaired individuals is poorer than that of the normal hearing individuals. This is probably due to the fact that they are unable to utilize more dynamic cues to consonant identity like formant transitions, and rely more on steady state spectral cues like burst frequency and frication energy, which being softer in intensity were not always available to them as the number of sine waves decreased.

Although it had been hypothesized that the sine wave modeling method would actually improve speech recognition performance when the original signal had a positive signal-to-noise ratio, this did not prove to be the case. When speech-in-noise was processed using sine wave modeling, peaks in the noise were presumably picked as speech peaks by the processing algorithm and reproduced in the sine wave versions. While consonant recognition in noise increased with an increasing number of sine waves, performance never approximated the performance on unprocessed speech.

Consonant recognition performance decreased with decreasing signal-to-noise ratio and was significantly poorer than the unprocessed speech in noise. The information transfer analysis revealed that as the signal-to-noise ratio decreased, the place information was lost first, followed by manner, and finally voicing cues. The spectrographic analysis revealed that the addition of noise obscured formant transitions, burst and frication energy. This would explain the reduction in consonant recognition performance in both groups. For the hearing-impaired listeners, poor spectral and temporal resolution may have contributed to their inability to use the few acoustic cues that were present in the sine wave modeled signal resulting in very poor consonant recognition scores in noise.

APPENDIX 1

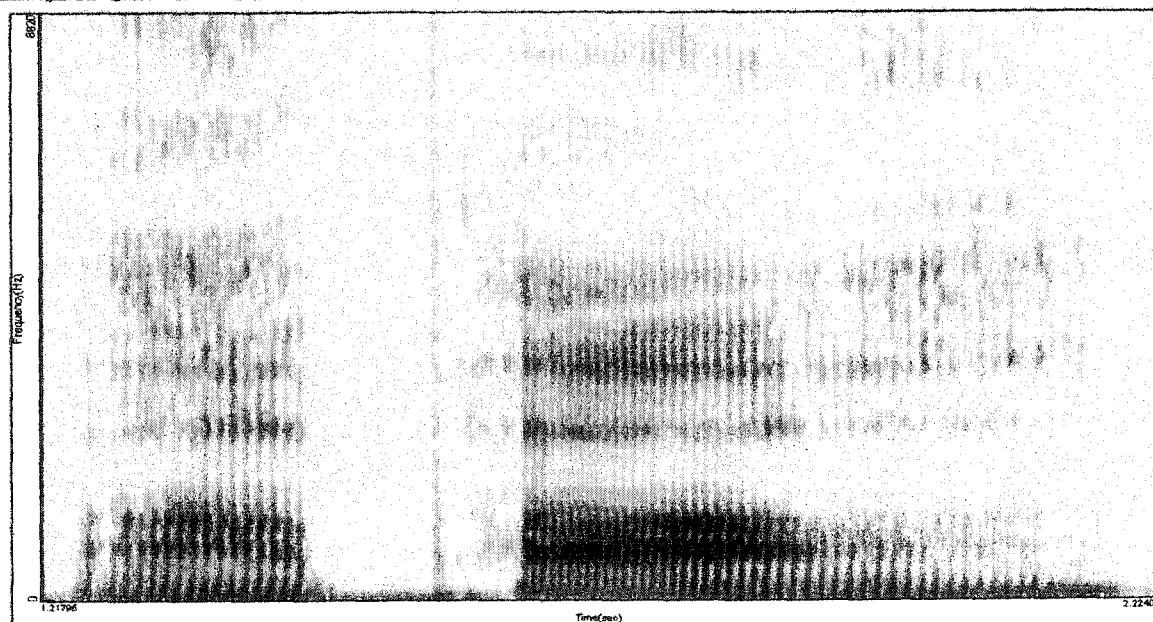
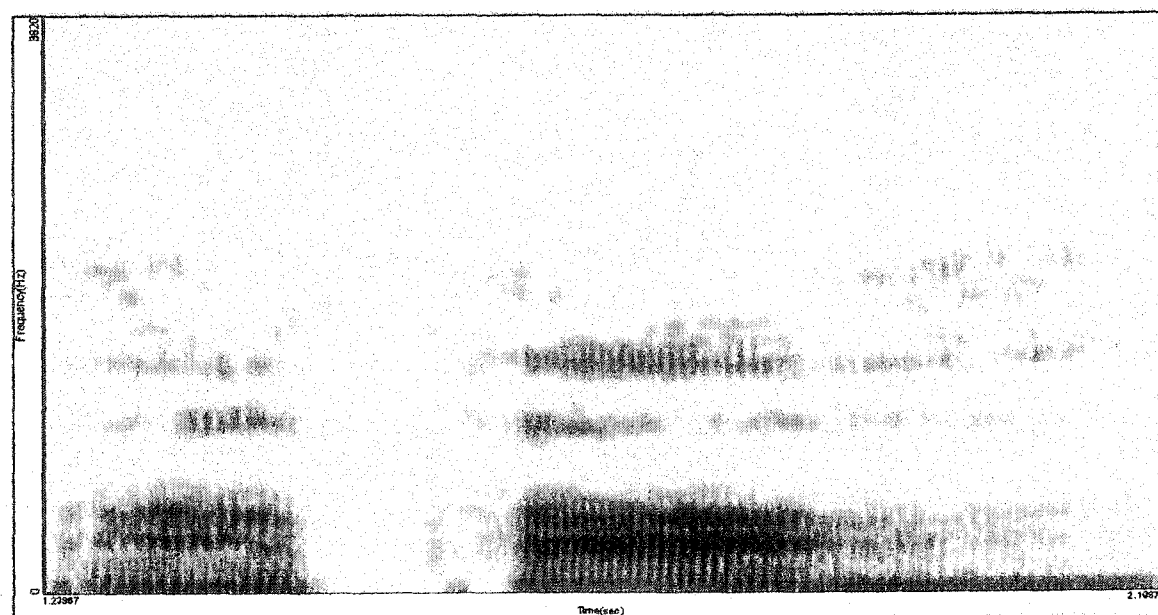
Wideband spectrograms of selected syllablesSyllable: /apa/Figure 1. Wideband spectrogram of /apa/ in quiet; Unprocessed stimuli
(Frequency scale: 0-8820 Hz).Figure 2. Wideband spectrogram of /apa/ in quiet; 16 sine waves
(Frequency scale: 0-8820 Hz).

Figure 3. Wideband spectrogram of /apa/ in quiet; 12 sine waves
(Frequency scale: 0-8820 Hz).

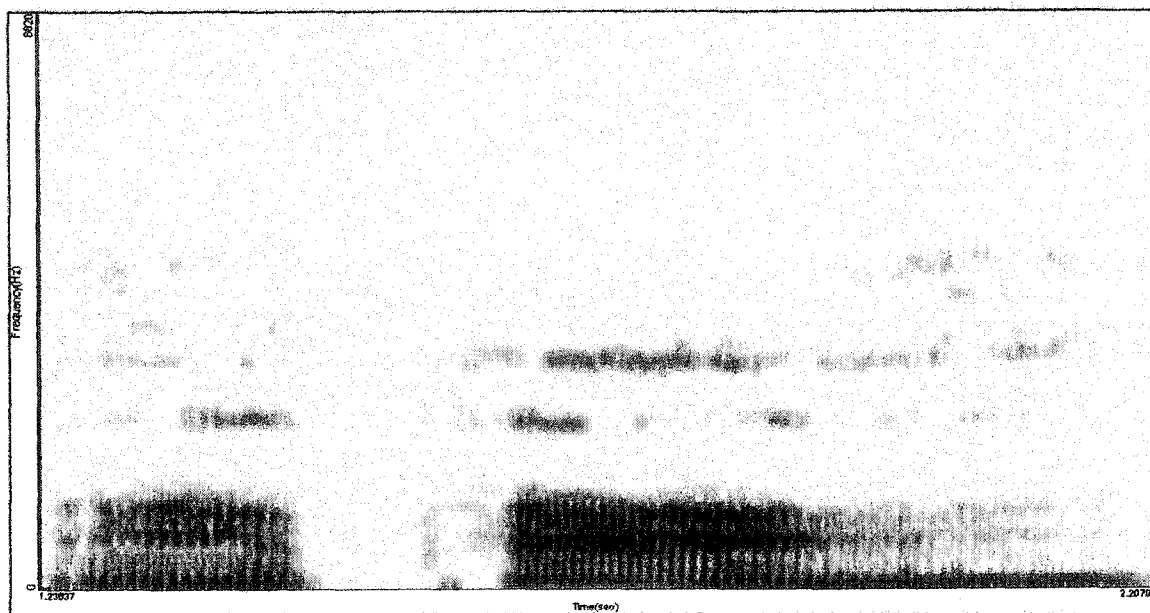


Figure 4. Wideband spectrogram of /apa/ in quiet; 8 sine waves
(Frequency scale: 0-8820 Hz).

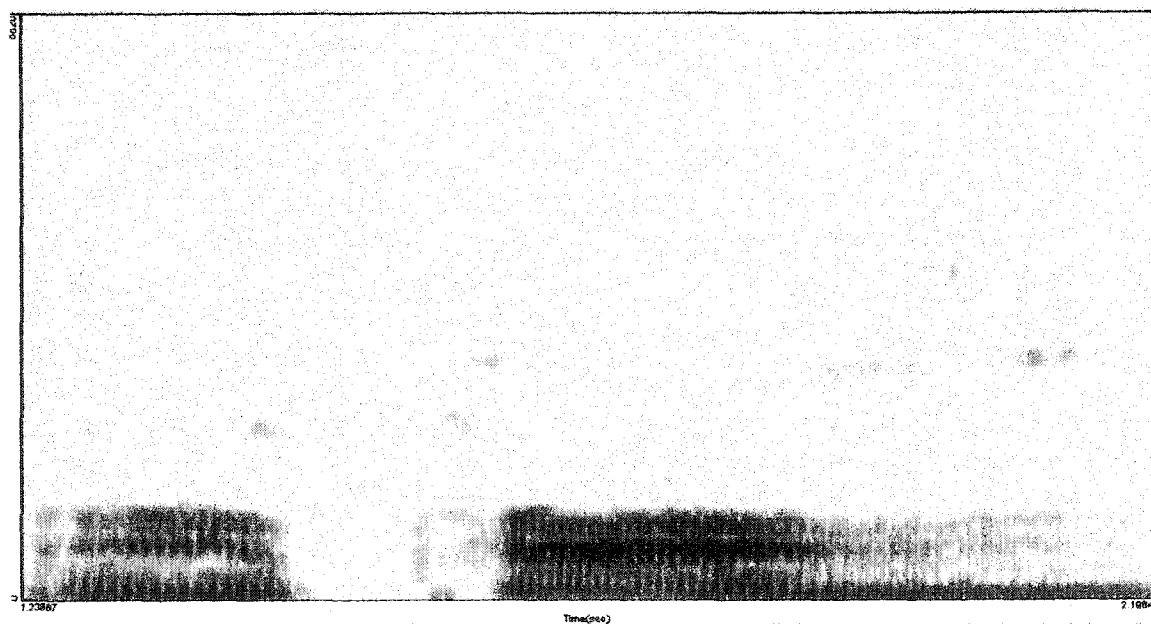


Figure 5. Wideband spectrogram of /apa/ in quiet; 4 sine waves
(Frequency scale: 0-8820 Hz).



Figure 6. Wideband spectrogram of /apa/ in +10SNR; unprocessed
(Frequency scale: 0-8820 Hz).

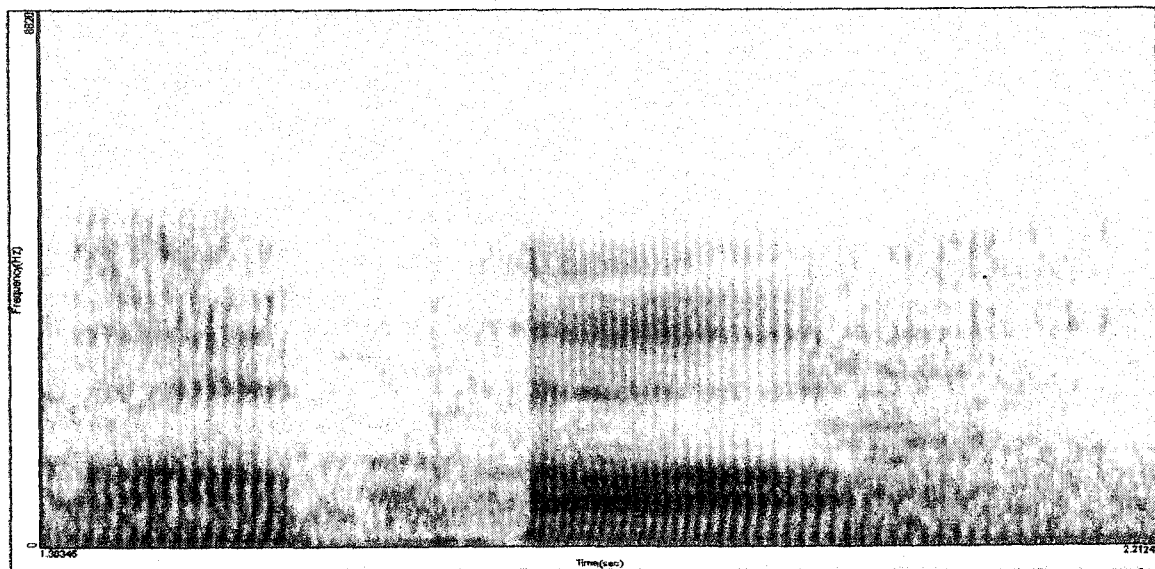


Figure 7. Wideband spectrogram of /apa/ in +10SNR; 16 sine waves
(Frequency scale: 0-8820 Hz).



Figure 8. Wideband spectrogram of /apa/ in +10SNR; 12 sine waves
(Frequency scale: 0-8820 Hz).



Figure 9. Wideband spectrogram of /apa/ in +10SNR; 8 sine waves
(Frequency scale: 0-8820 Hz).

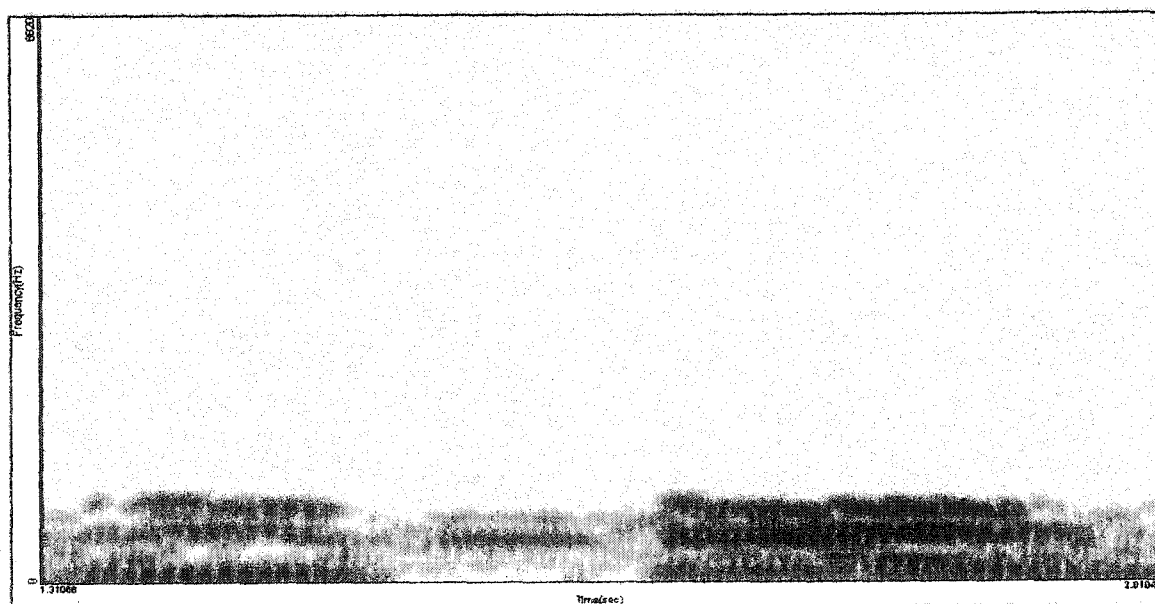


Figure 10. Wideband spectrogram of /apa/ in +10SNR; 4 sine waves
(Frequency scale: 0-8820 Hz).

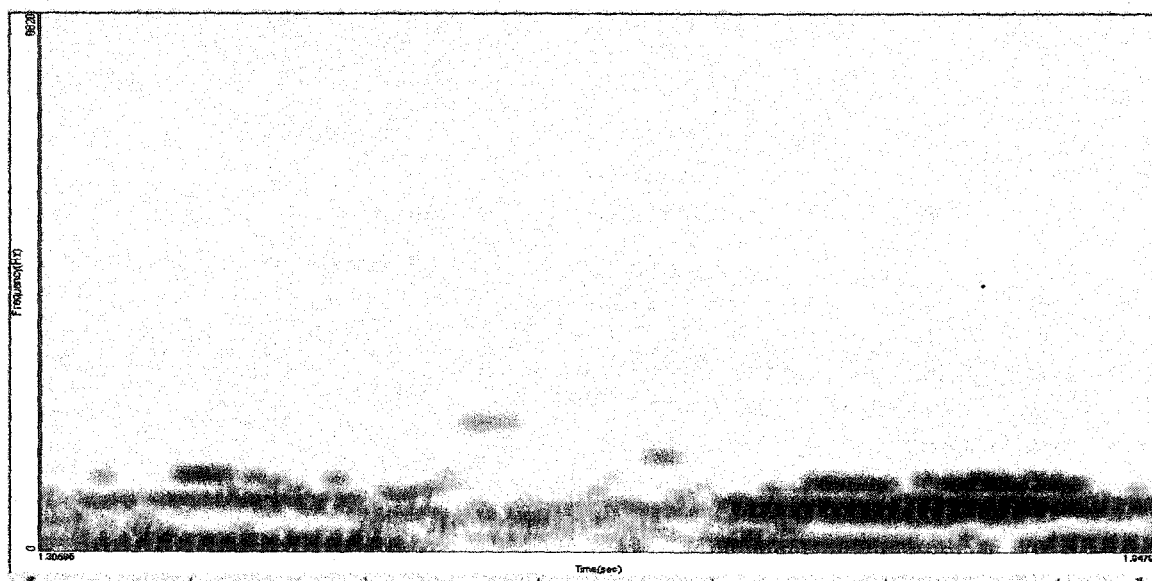


Figure 11. Wideband spectrogram of /apa/ in 0SNR; unprocessed
(Frequency scale: 0-11205 Hz).



Figure 12. Wideband spectrogram of /apa/ in 0SNR; 16 sine waves
(Frequency scale: 0-11205 Hz).



Figure 13. Wideband spectrogram of /apa/ in 0SNR; 12 sine waves
(Frequency scale: 0-11205 Hz).

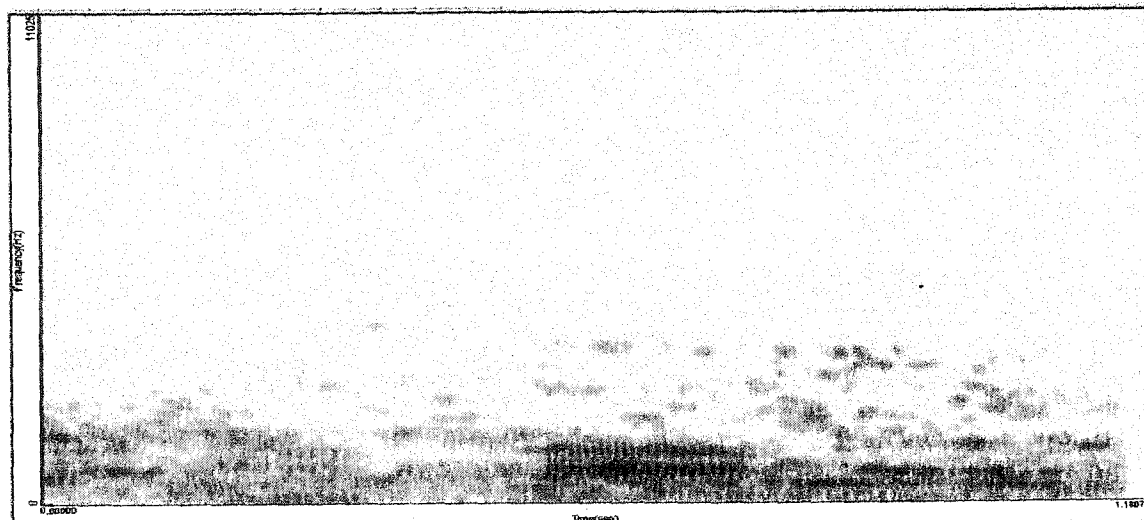
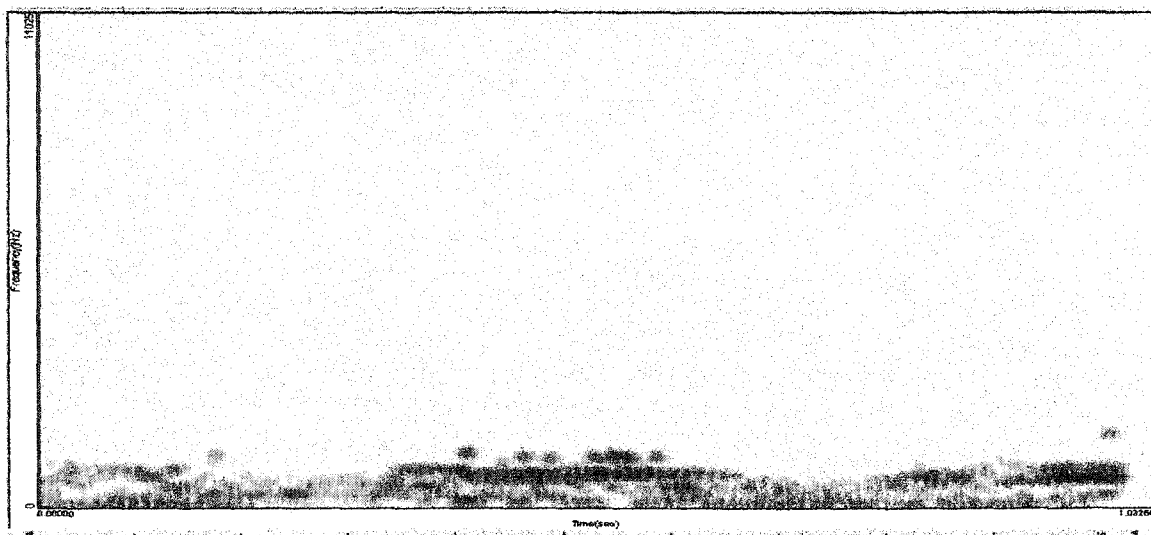


Figure 14. Wideband spectrogram of /apa/ in 0SNR; 8 sine waves
(Frequency scale: 0-11205 Hz).



Figure 15. Wideband spectrogram of /apa/ in 0SNR; 4 sine waves
(Frequency scale: 0-11205 Hz).



Syllable: /aba/

Figure 16. Wideband spectrogram of /aba/ in quiet; Unprocessed stimuli
(Frequency scale: 0-8820 Hz).

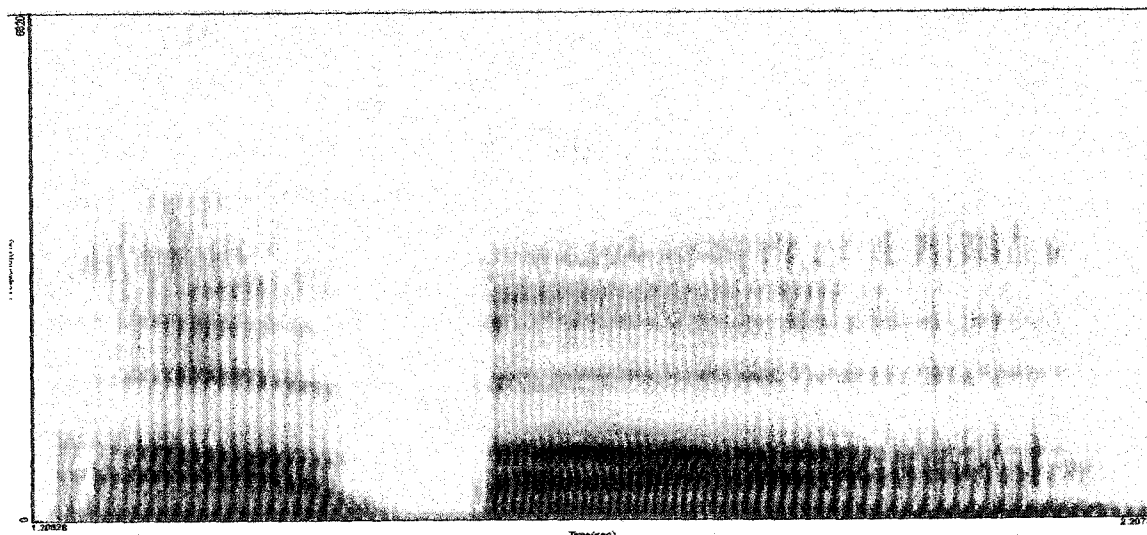


Figure 17. Wideband spectrogram of /aba/ in quiet; 16sine waves
(Frequency scale: 0-8820 Hz).

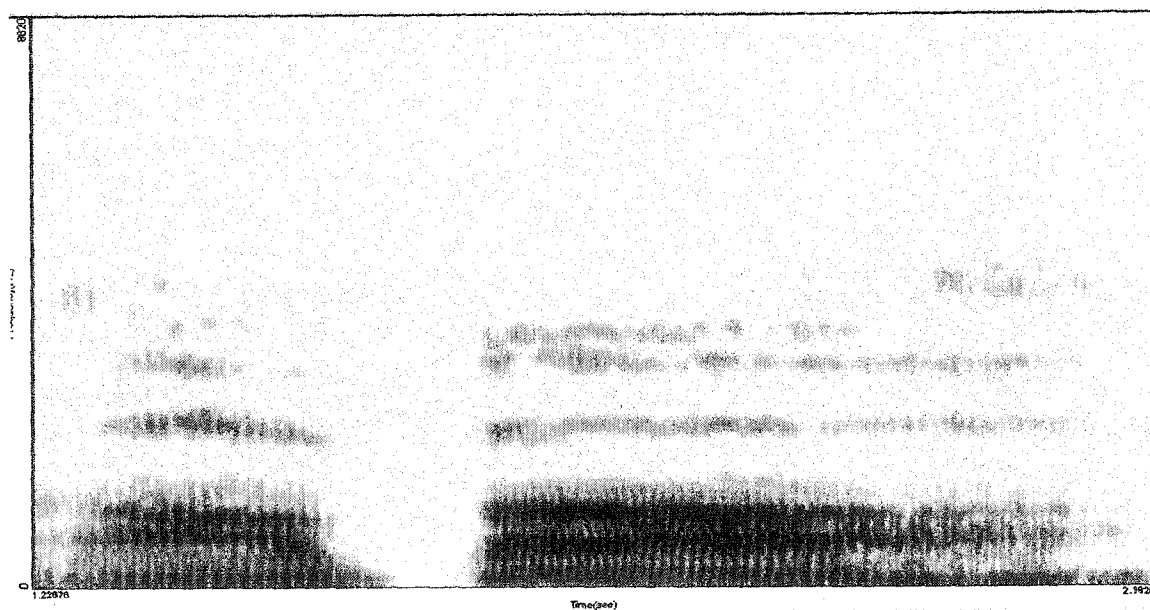


Figure 18. Wideband spectrogram of /aba/ in quiet; 12 sine waves
(Frequency scale: 0-8820 Hz).

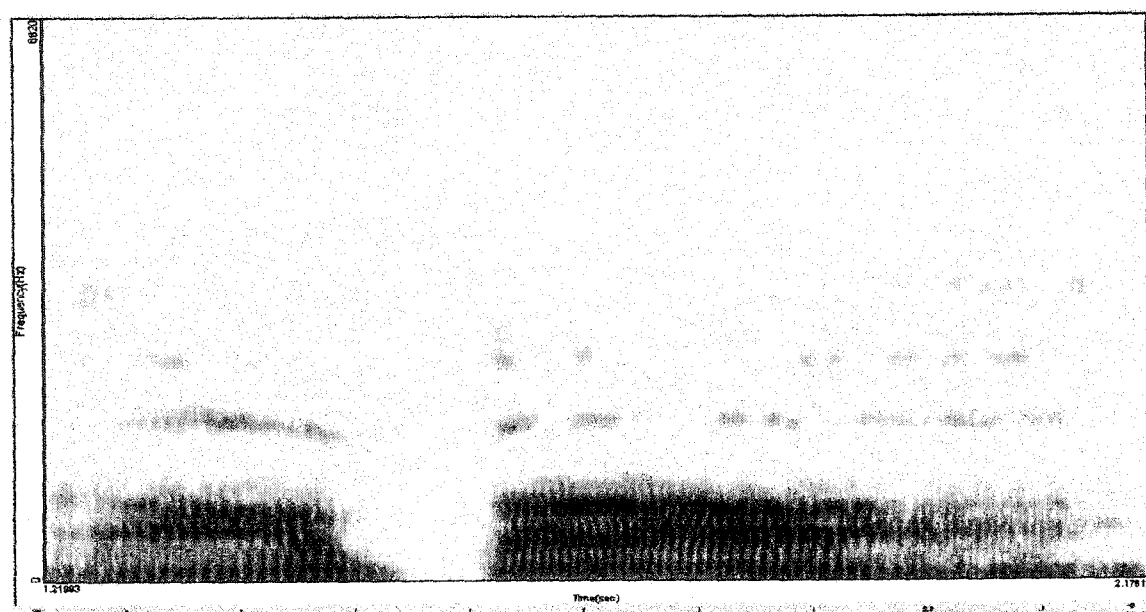


Figure 19. Wideband spectrogram of /aba/ in quiet; 8 sine waves.
(Frequency scale: 0-8820 Hz).



Figure 20. Wideband spectrogram of /aba/ in quiet; 4 sine waves
(Frequency scale: 0-8820 Hz).

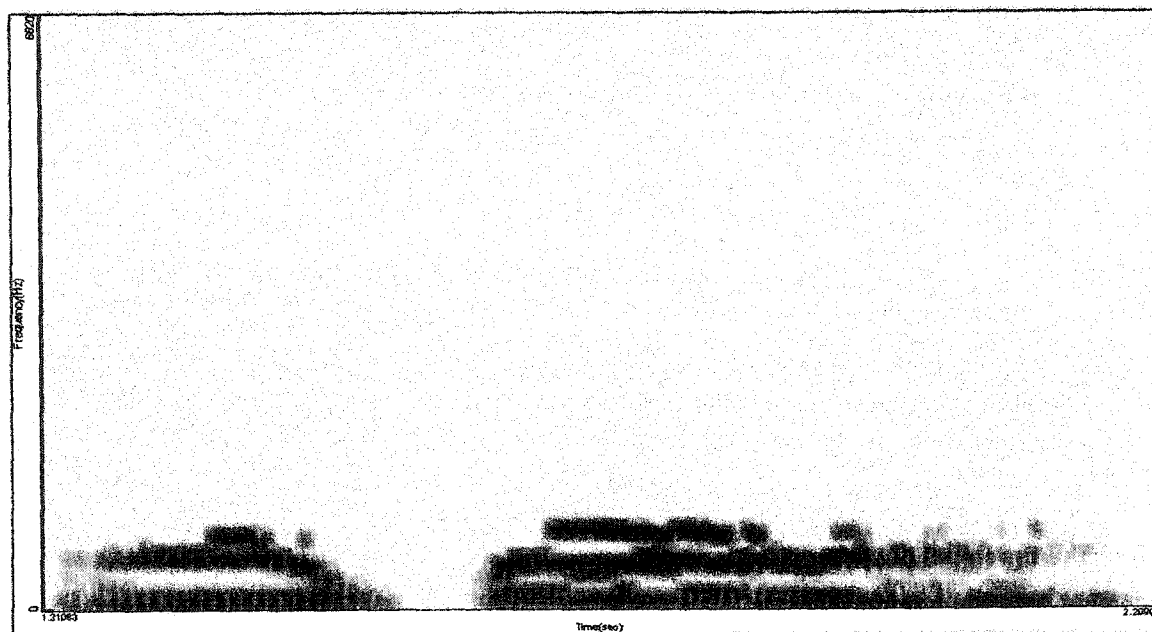


Figure 21. Wideband spectrogram of /aba/ in +10SNR; Unprocessed
(Frequency scale: 0-8820 Hz).



Figure 22. Wideband spectrogram of /aba/ in +10SNR; 16sine waves
(Frequency scale: 0-8820 Hz).

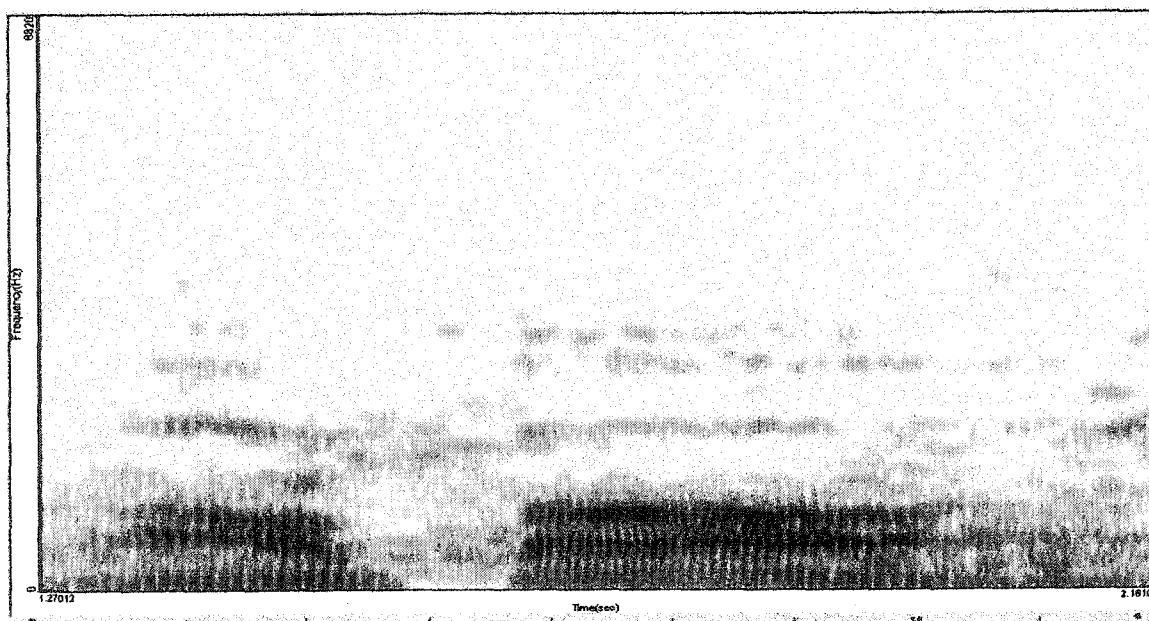


Figure 23. Wideband spectrogram of /aba/ in +10SNR; 12 sine waves
 (Frequency scale: 0-8820 Hz).

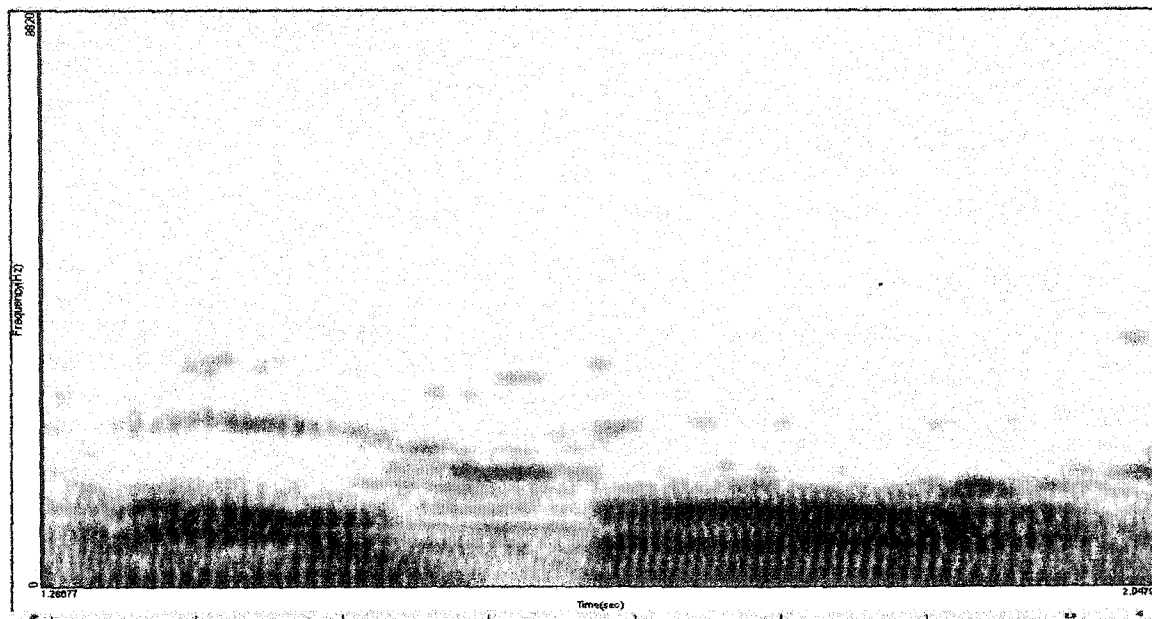


Figure 24. Wideband spectrogram of /aba/ in +10SNR; 8 sine waves
 (Frequency scale: 0-8820 Hz).

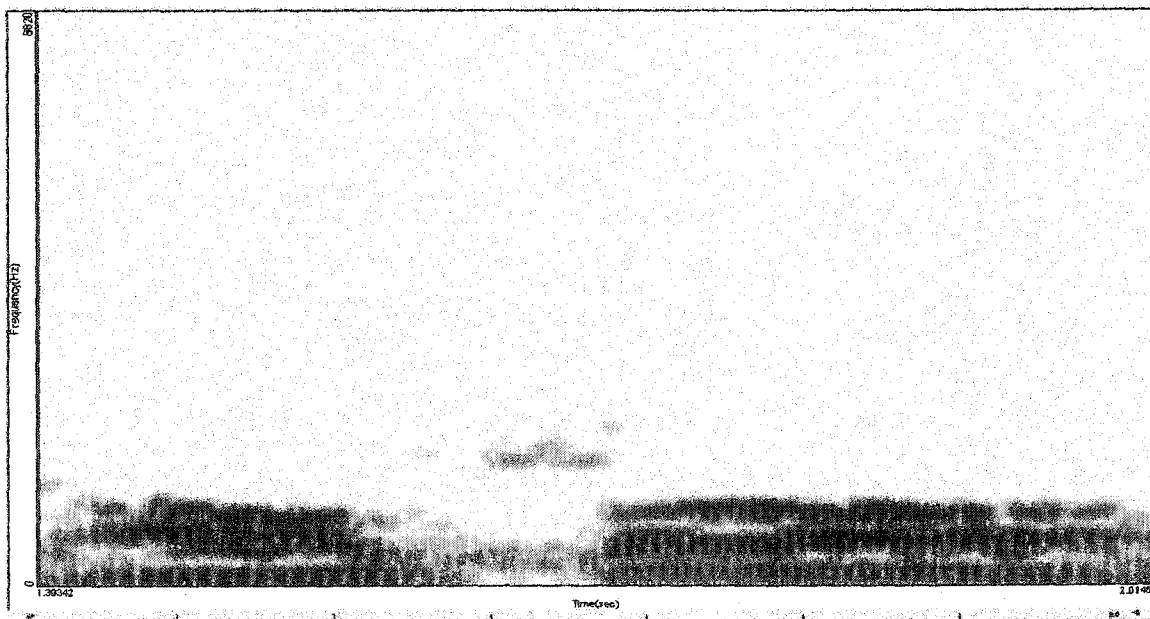


Figure 24. Wideband spectrogram of /aba/ in +10SNR; 4 sine waves.
(Frequency scale: 0-8820 Hz)

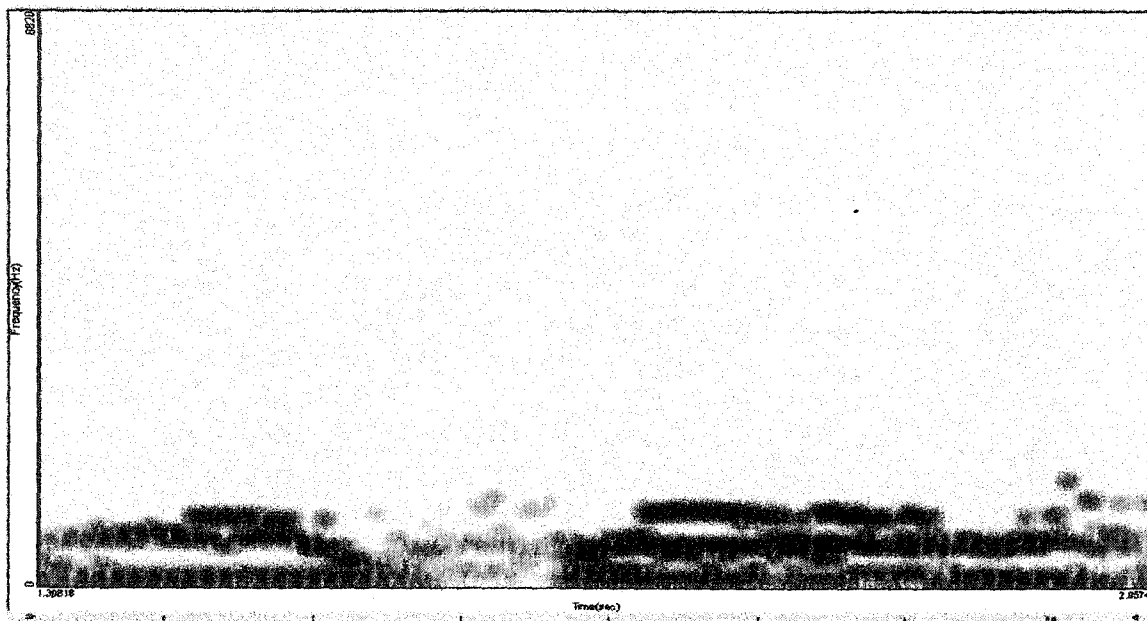


Figure 26. Wideband spectrogram of /aba/ in 0SNR; unprocessed
(Frequency scale: 0-11025 Hz).

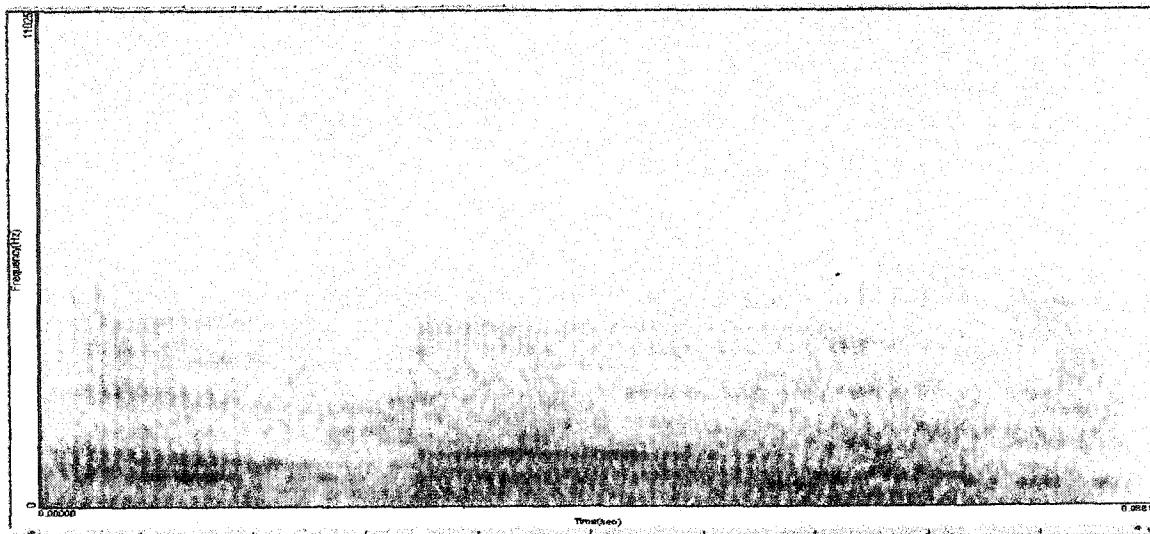


Figure 27. Wideband spectrogram of /aba/ in 0SNR; 16 sine waves
(Frequency scale: 0-11025 Hz).

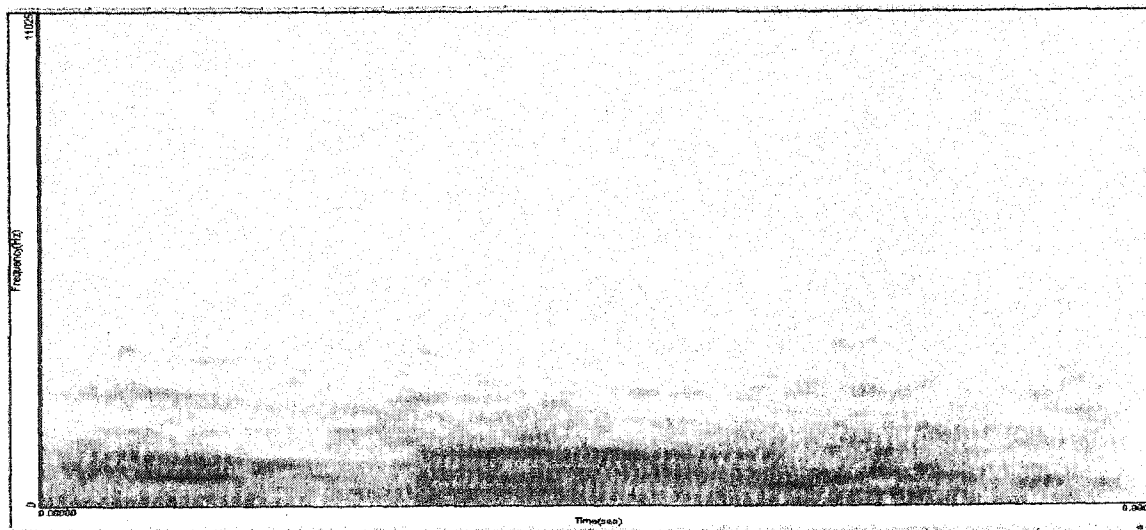


Figure 28. Wideband spectrogram of /aba/ in 0SNR; 12 sine waves
(Frequency scale: 0-11025 Hz).



Figure 29. Wideband spectrogram of /aba/ in 0SNR; 8 sine waves
(Frequency scale: 0-11025 Hz).

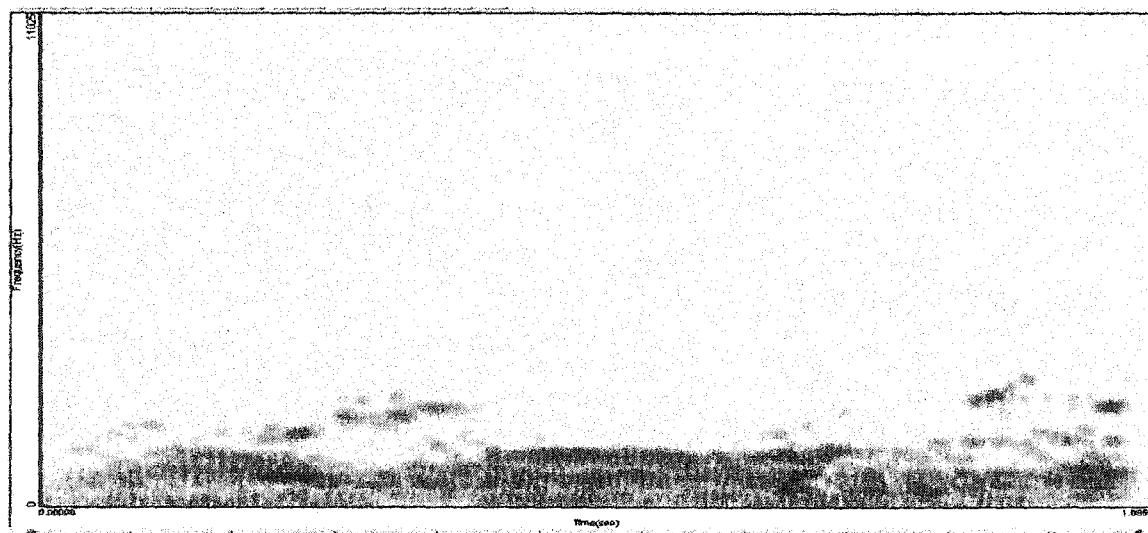


Figure 30. Wideband spectrogram of /aba/ in 0SNR; 4 sine waves
(Frequency scale: 0-11025 Hz).



Syllable: /asa/

Figure 31. Wideband spectrogram of /asa/ in Quiet; unprocessed
(Frequency scale: 0-8820 Hz).

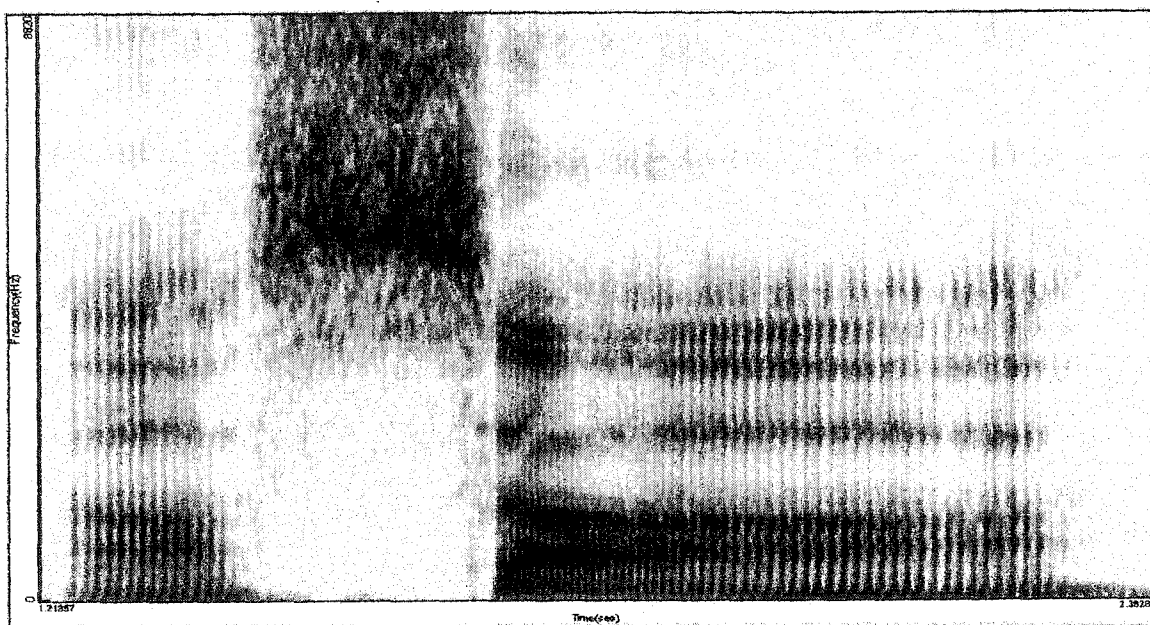


Figure 32. Wideband spectrogram of /asa/ in Quiet; 16 sine waves
(Frequency scale: 0-8820 Hz).



Figure 33. Wideband spectrogram of /asa/ in Quiet; 12 sine waves
(Frequency scale: 0-8820 Hz).

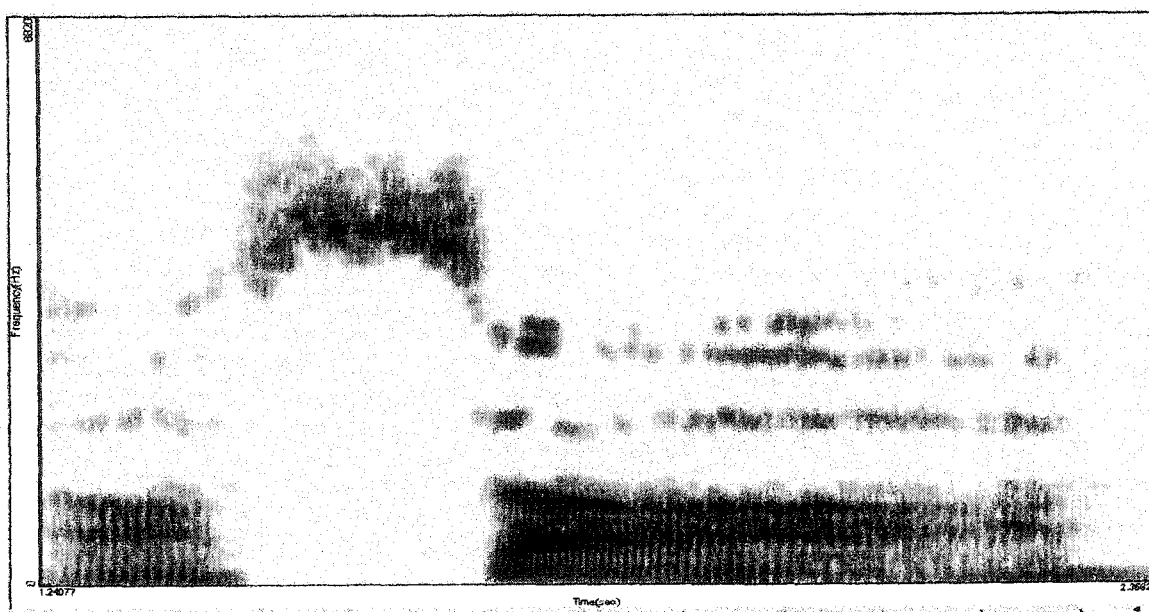


Figure 34. Wideband spectrogram of /asa/ in Quiet; 8 sine waves
(Frequency scale: 0-8820 Hz).

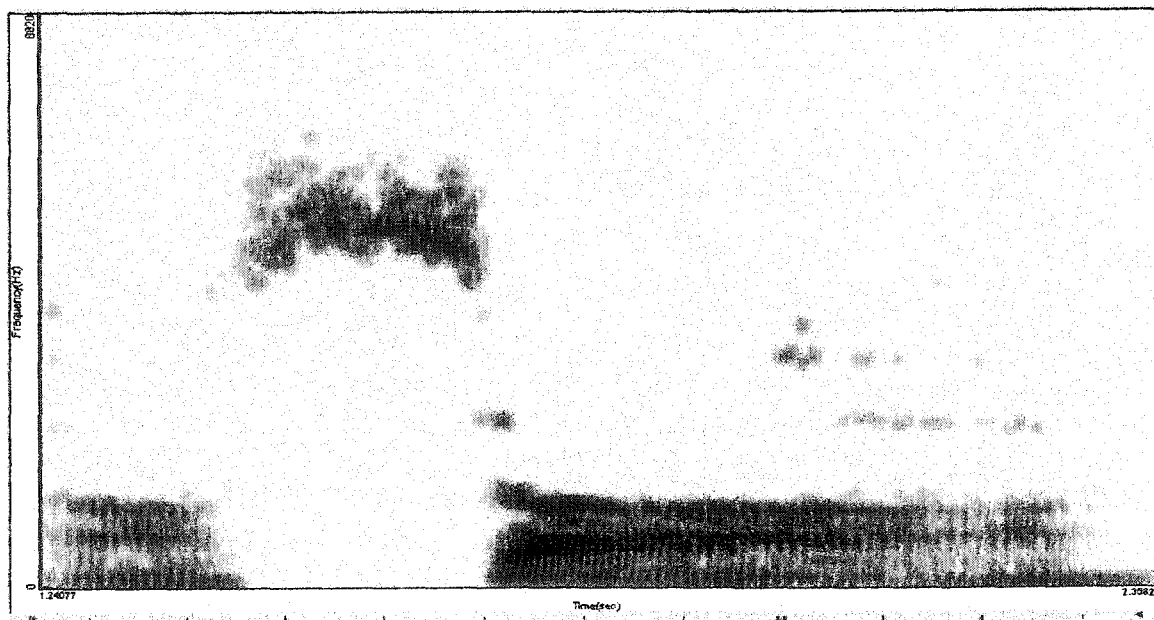


Figure 35. Wideband spectrogram of /asa/ in Quiet; 4 sine waves
(Frequency scale: 0-8820 Hz)

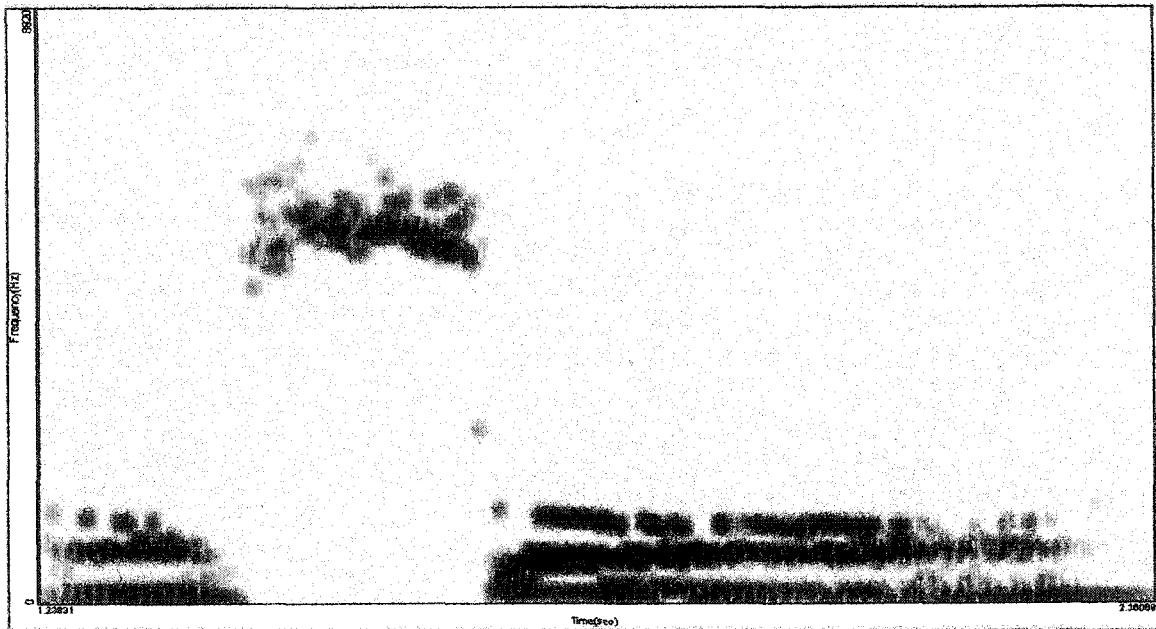


Figure 36. Wideband spectrogram of /asa/ in +10SNR; unprocessed
 (Frequency scale: 0-8820 Hz)

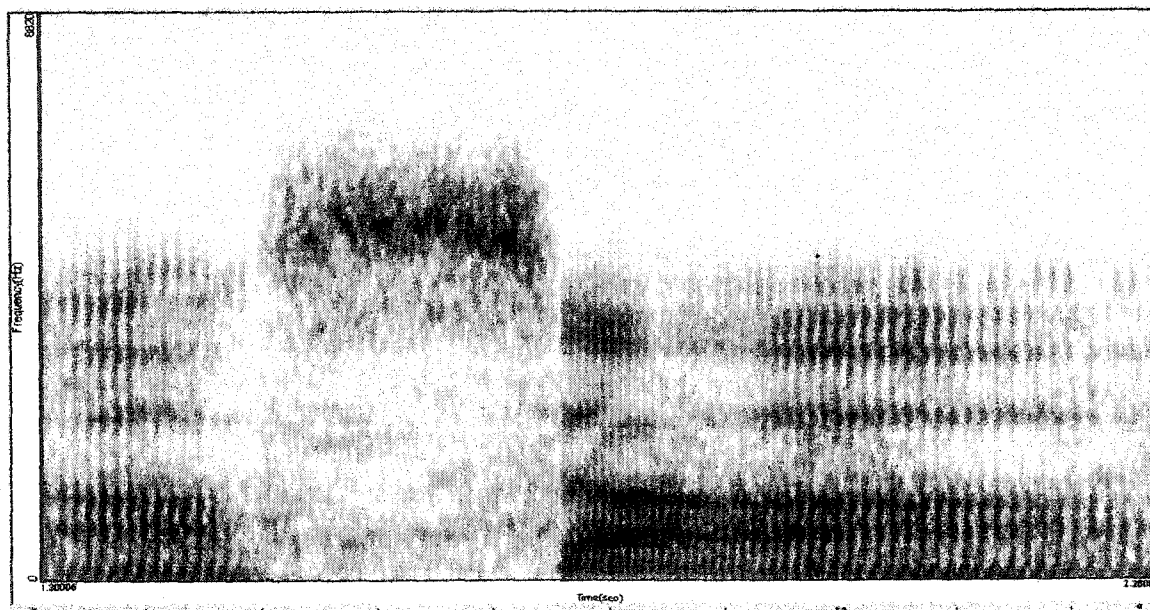


Figure 37. Wideband spectrogram of /asa/ in +10SNR; 16 sine waves
 (Frequency scale: 0-8820 Hz)



Figure 38. Wideband spectrogram of /asa/ in +10SNR; 12 sine waves
(Frequency scale: 0-8820 Hz)



Figure 39. Wideband spectrogram of /asa/ in +10SNR; 8 sine waves
(Frequency scale: 0-8820 Hz)

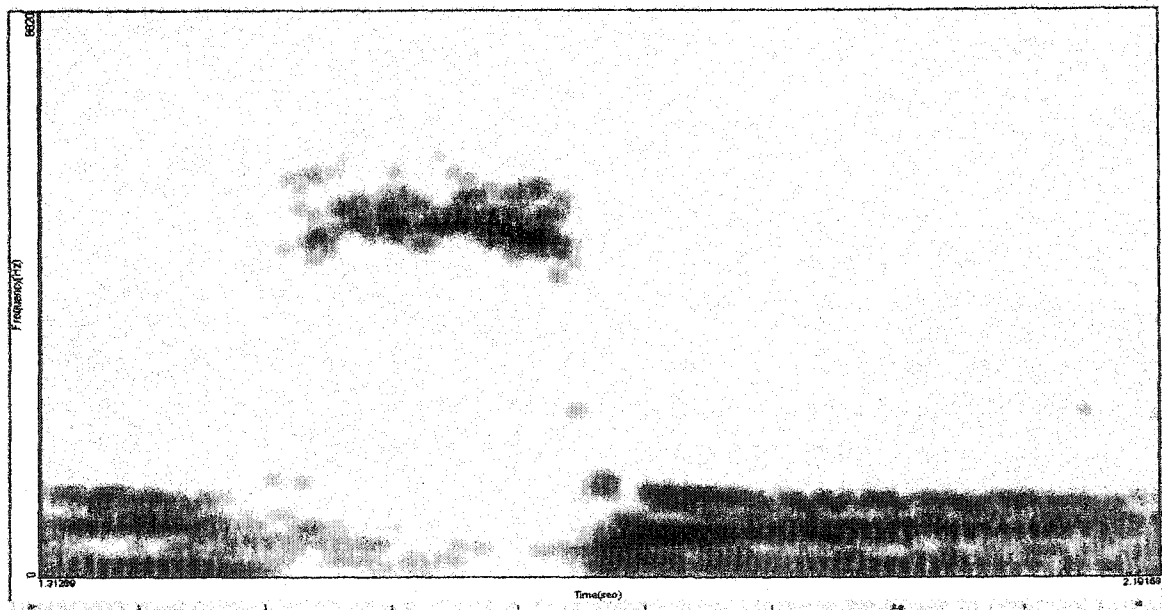


Figure 40. Wideband spectrogram of /asa/ in +10SNR; 4 sine waves
(Frequency scale: 0-8820 Hz)



Figure 41. Wideband spectrogram of /asa/ in 0SNR; unprocessed
(Frequency scale: 0-11025 Hz).

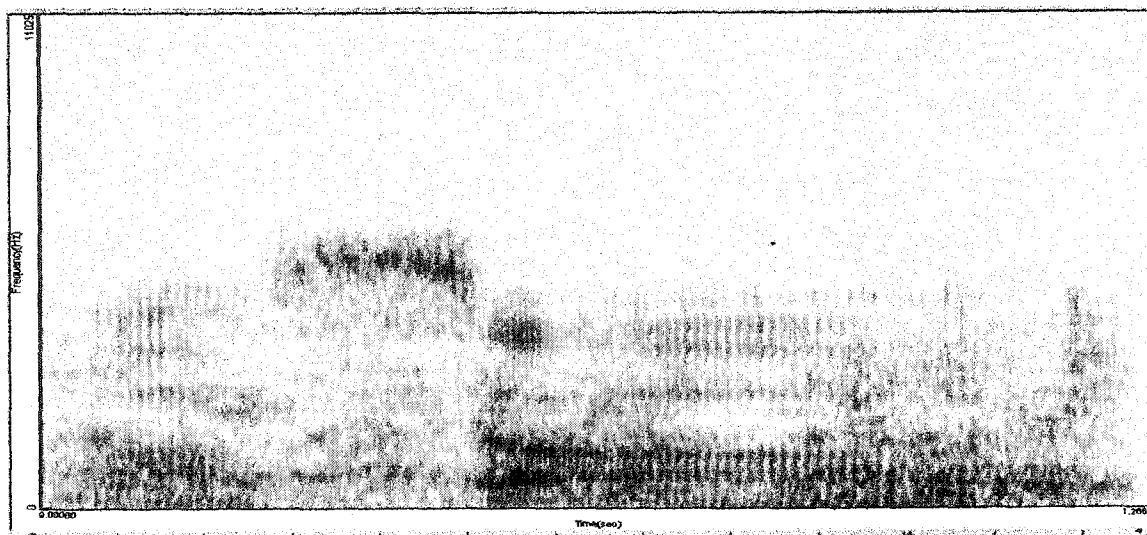


Figure 42. Wideband spectrogram of /asa/ in 0SNR; 16 sine waves
(Frequency scale: 0-11025 Hz).



Figure 43. Wideband spectrogram of /asa/ in 0SNR; 12 sine waves
(Frequency scale: 0-11025 Hz).



Figure 44. Wideband spectrogram of /asa/ in 0SNR; 8 sine waves
(Frequency scale: 0-11025 Hz).



Figure 45. Wideband spectrogram of /asa/ in 0SNR; 4 sine waves
(Frequency scale: 0-11025 Hz).



Syllable /aza/

Figure 46. Wideband spectrogram of /aza/ in Quiet; unprocessed
(Frequency scale: 0-8820 Hz)

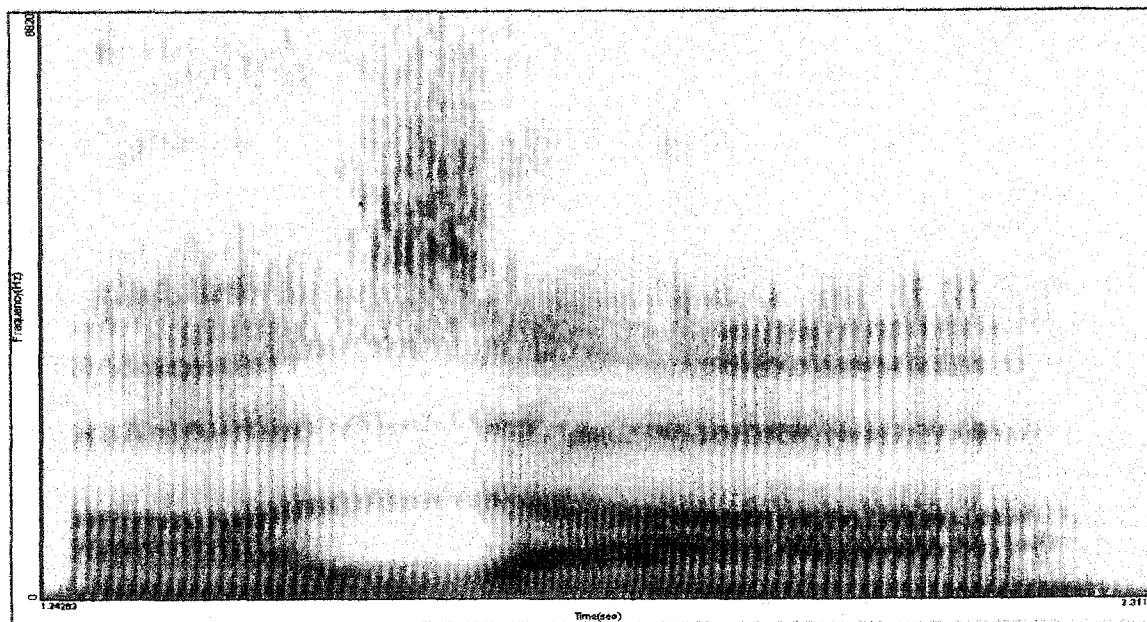


Figure 47. Wideband spectrogram of /aza/ in Quiet; 16 sine waves
(Frequency scale: 0-8820 Hz)

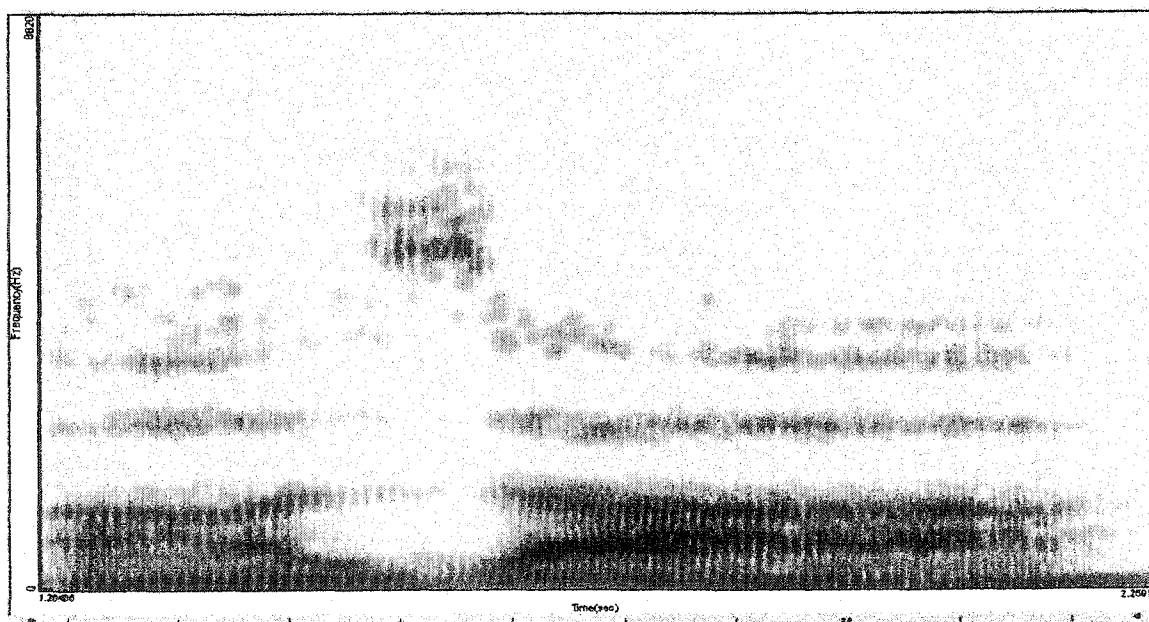


Figure 48. Wideband spectrogram of /aza/ in Quiet; 12 sine waves
(Frequency scale: 0-8820 Hz)

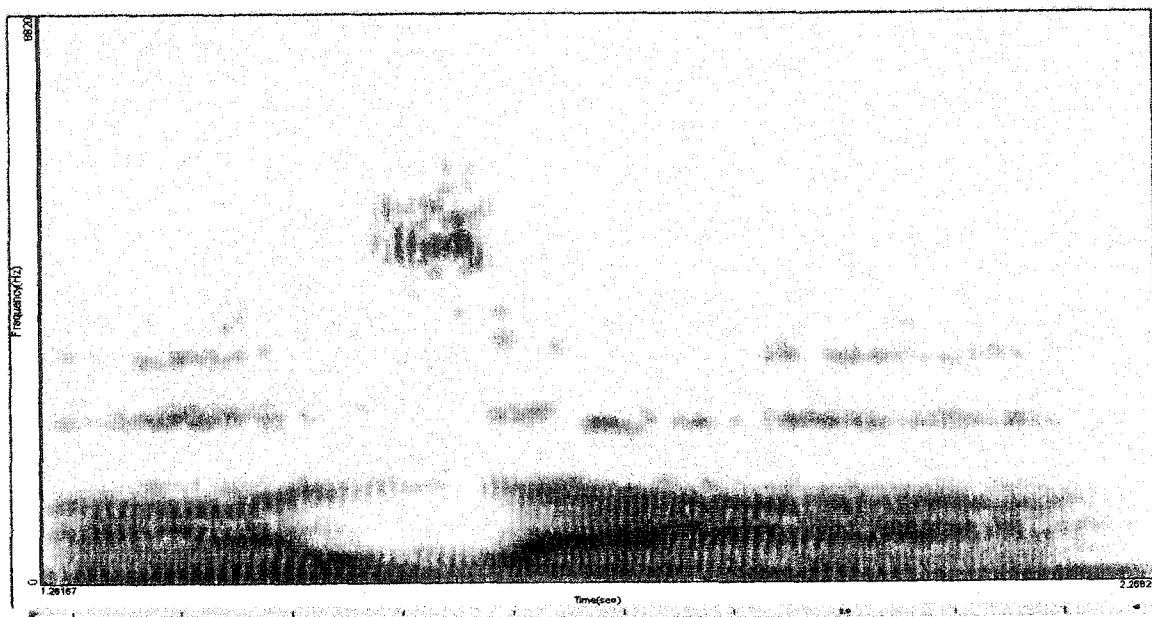


Figure 49. Wideband spectrogram of /aza/ in Quiet; 8 sine waves
(Frequency scale: 0-8820 Hz)

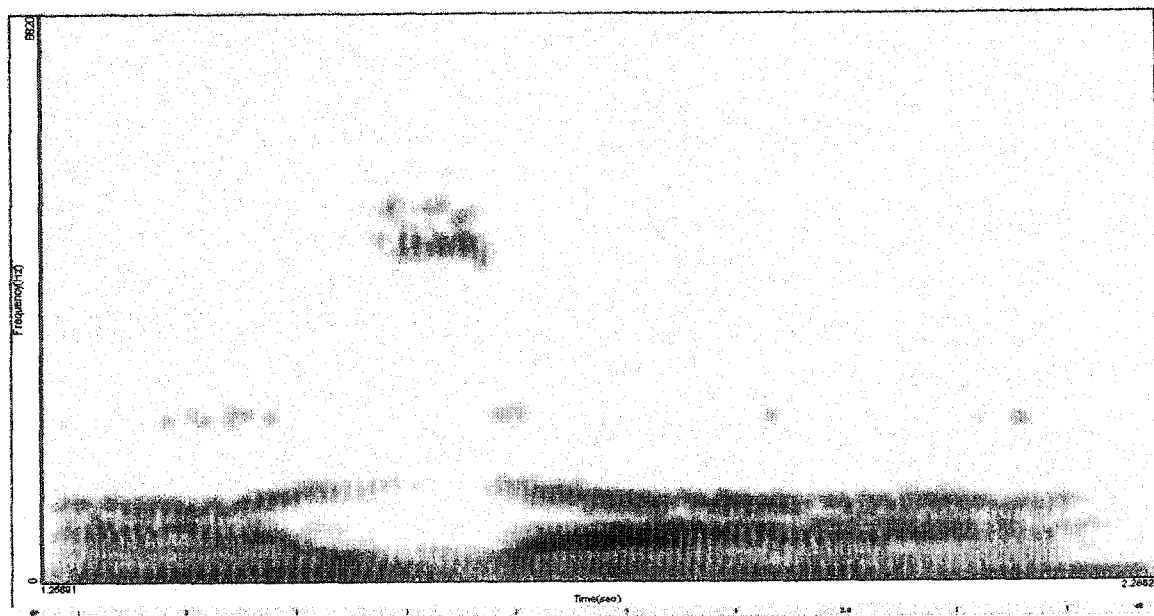


Figure 50. Wideband spectrogram of /aza/ in Quiet; 4 sine waves
(Frequency scale: 0-8820 Hz)

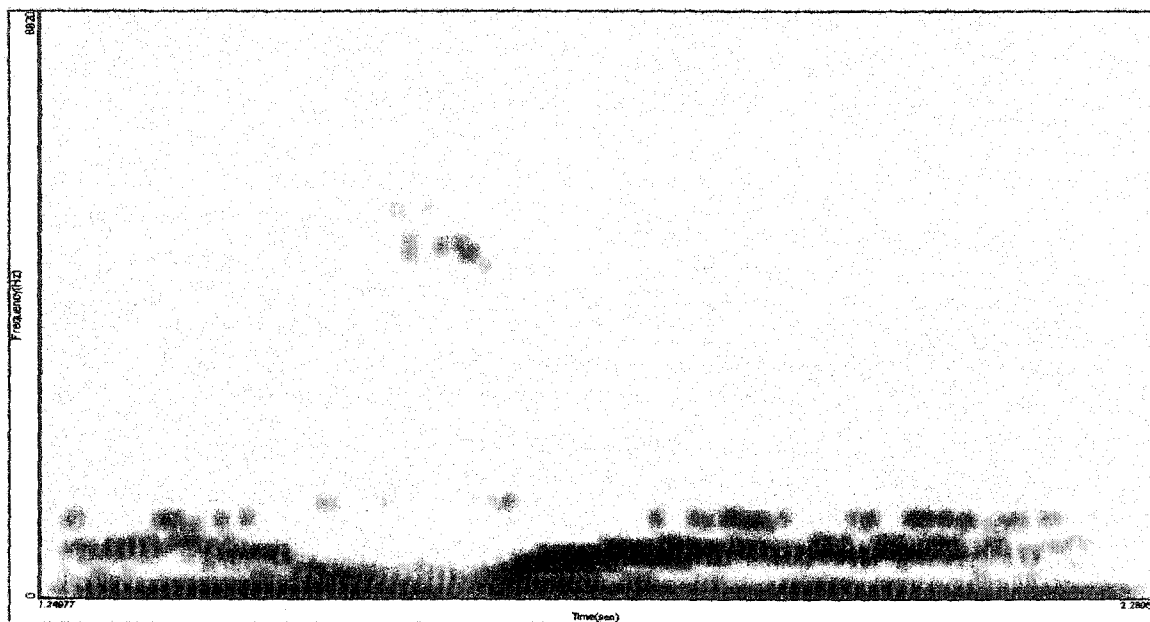


Figure 51. Wideband spectrogram of /aza/ in +10SNR; Unprocessed
(Frequency scale: 0-8820 Hz)

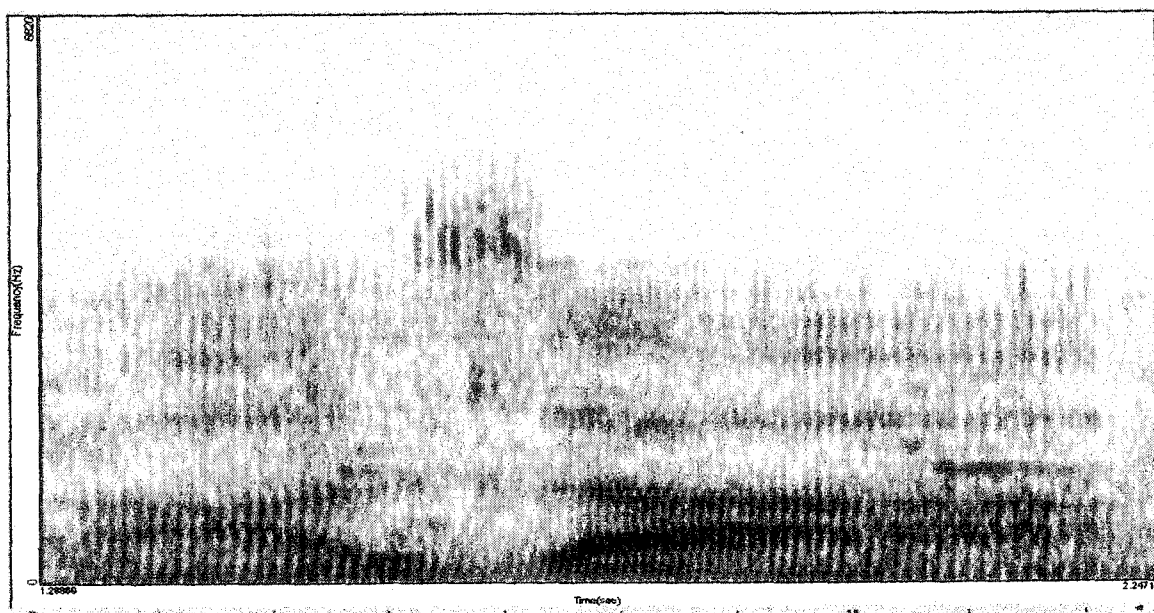


Figure 52. Wideband spectrogram of /aza/ in +10SNR; 16 sine waves
(Frequency scale: 0-8820 Hz)

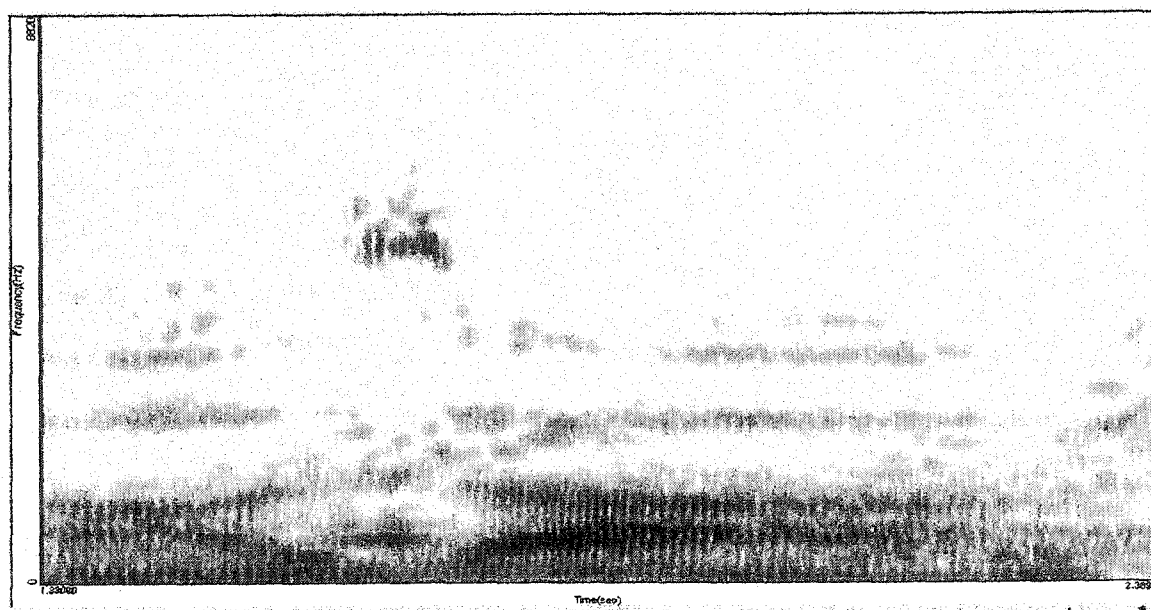


Figure 53. Wideband spectrogram of /aza/ in +10SNR; 12 sine waves
(Frequency scale: 0-8820 Hz)



Figure 54. Wideband spectrogram of /aza/ in +10SNR; 8 sine waves
(Frequency scale: 0-8820 Hz)

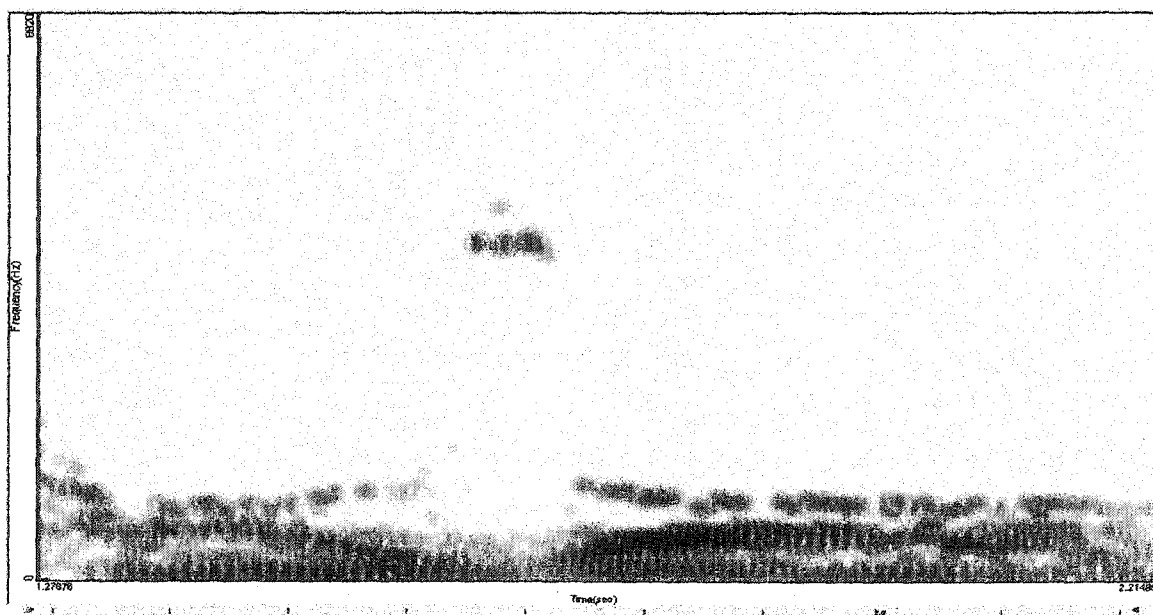


Figure 55. Wideband spectrogram of /aza/ in +10SNR; 4 sine waves
(Frequency scale: 0-8820 Hz)

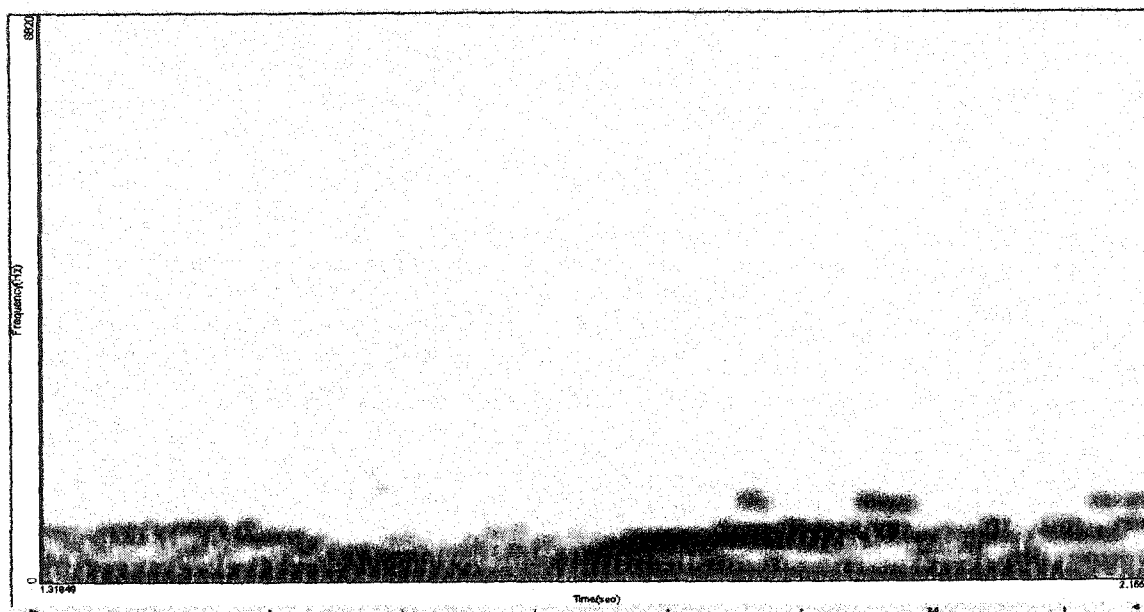


Figure 56. Wideband spectrogram of /aza/ in 0SNR; Unprocessed
(Frequency scale: 0-11025 Hz)

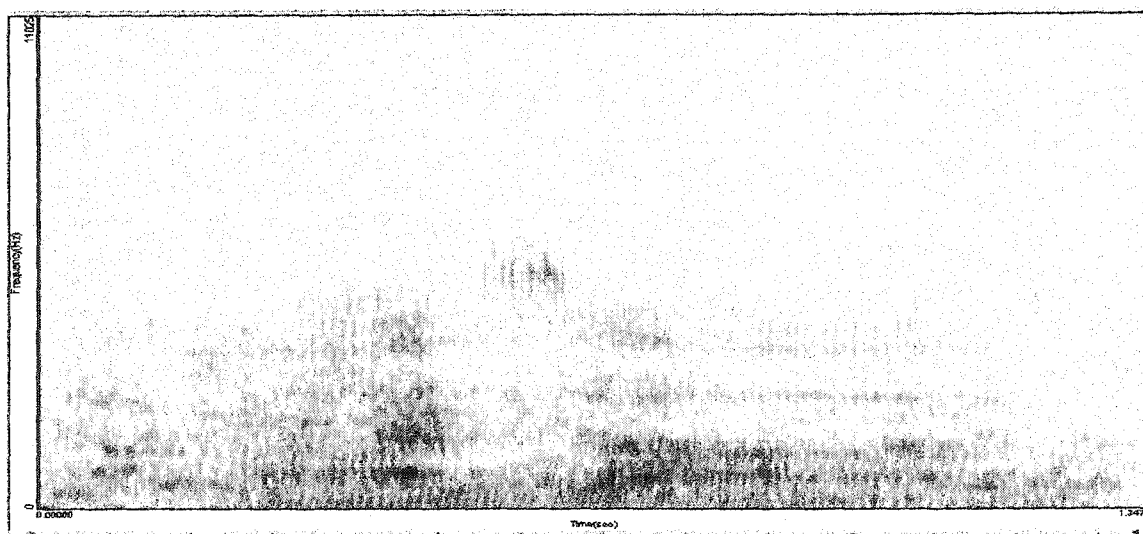


Figure 57. Wideband spectrogram of /aza/ in 0SNR; 16 sine waves
(Frequency scale: 0-11025 Hz).

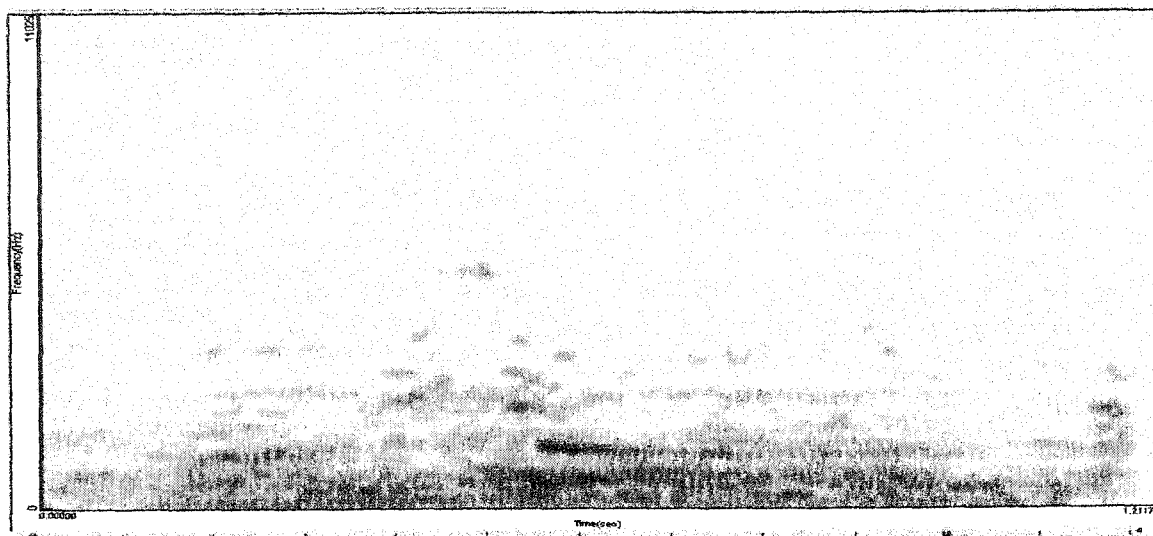


Figure 58. Wideband spectrogram of /aza/ in 0SNR; 12 sine waves
(Frequency scale: 0-11025 Hz).

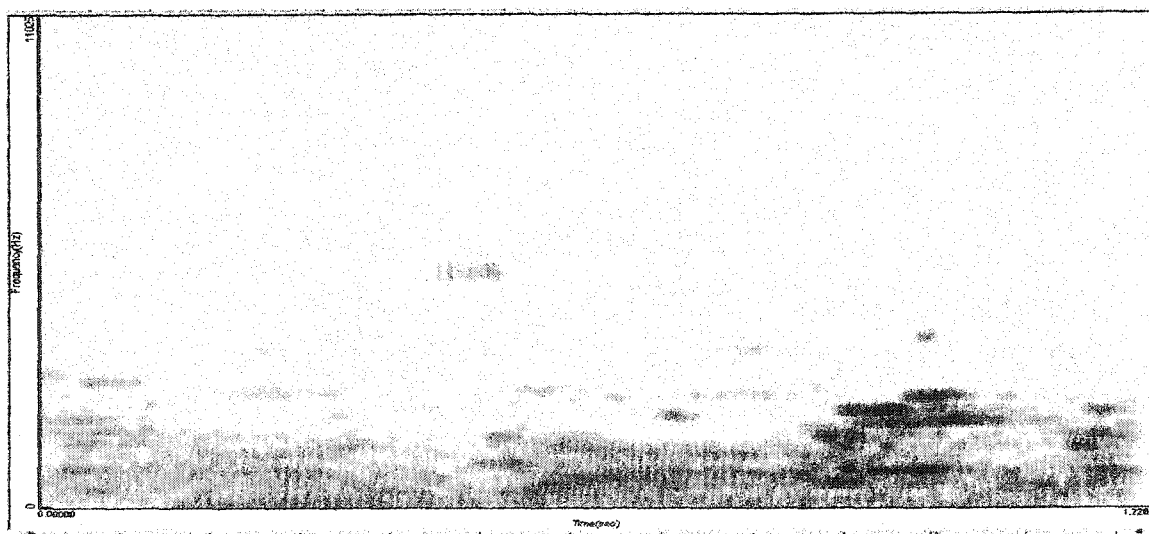


Figure 59. Wideband spectrogram of /aza/ in 0SNR; 8 sine waves
(Frequency scale: 0-11025 Hz).

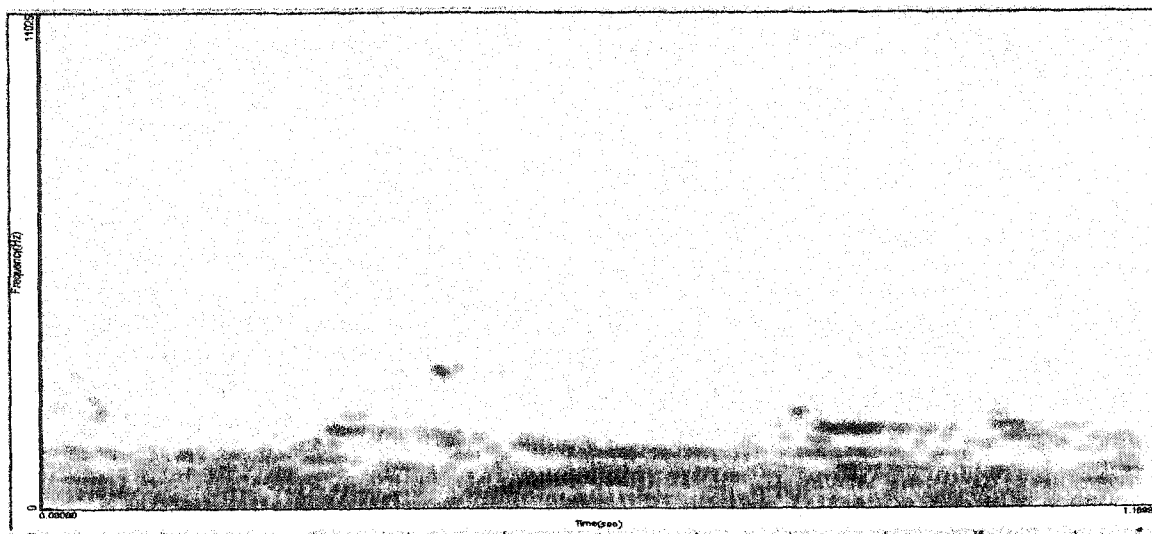
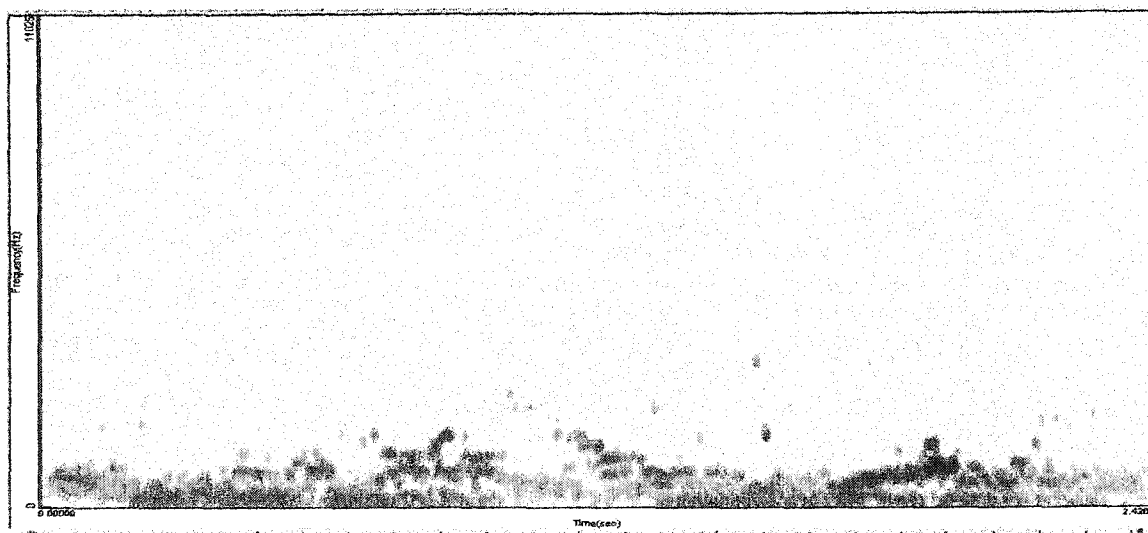


Figure 60. Wideband spectrogram of /aza/ in 0SNR; 4 sine waves
(Frequency scale: 0-11025 Hz).



Syllable /atʃa/

Figure 61. Wideband spectrogram of /atʃa/ in Quiet; Unprocessed
(Frequency scale: 0-8820 Hz).

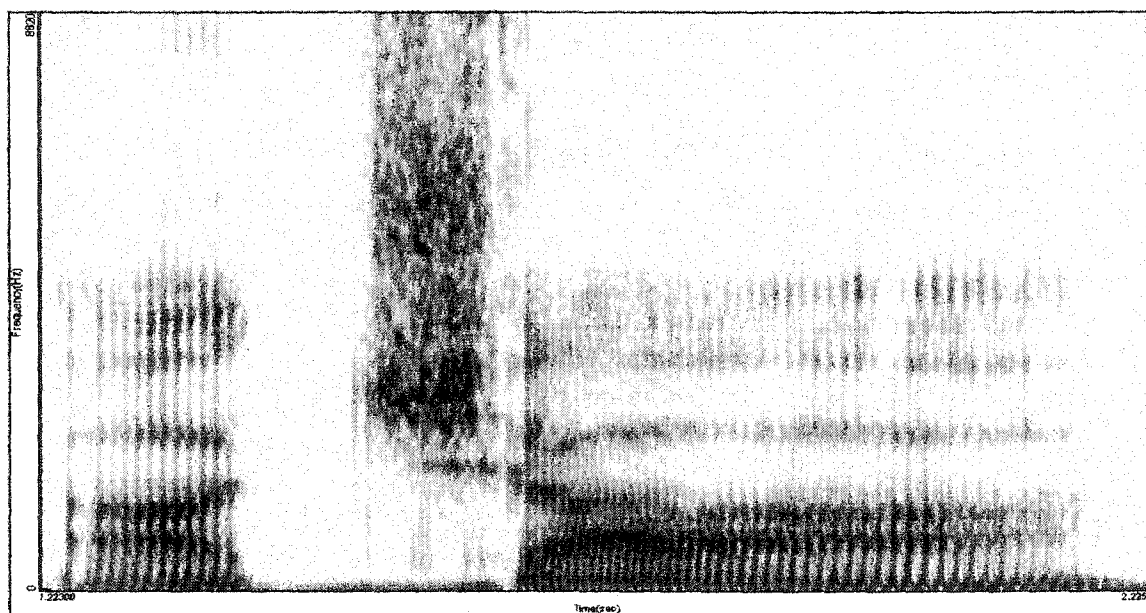


Figure 62. Wideband spectrogram of /atʃa/ in Quiet; 16 sine waves
(Frequency scale: 0-8820 Hz).



Figure 63. Wideband spectrogram of / atʃa / in Quiet; 12 sine waves
(Frequency scale: 0-8820 Hz).



Figure 64. Wideband spectrogram of / atʃa / in Quiet; 8 sine waves
(Frequency scale: 0-8820 Hz).

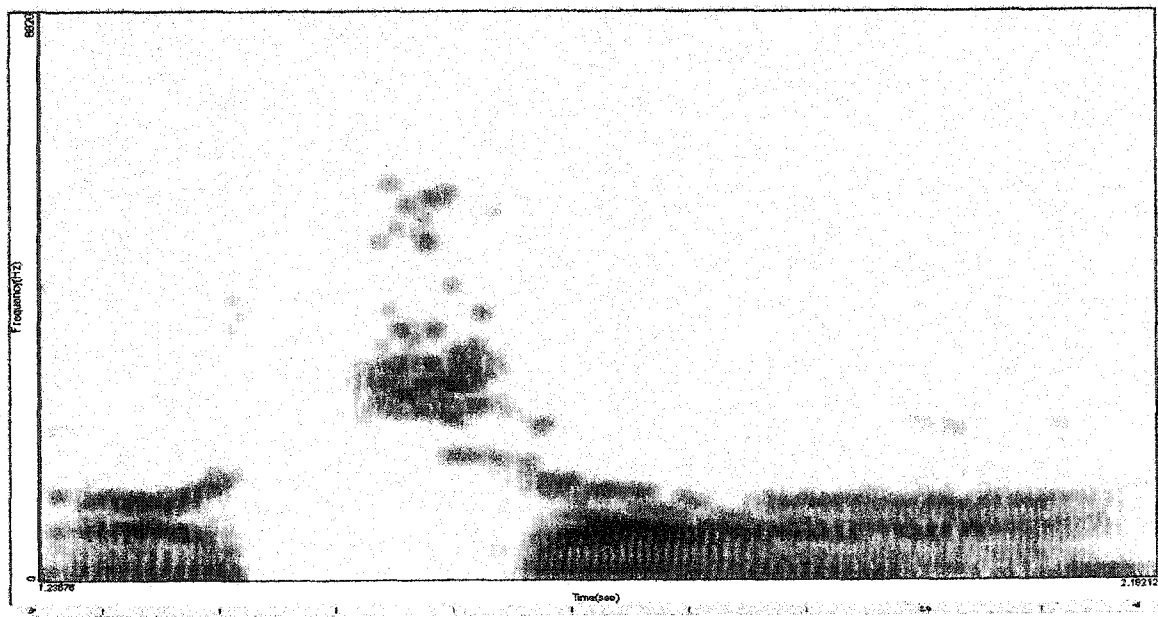


Figure 65. Wideband spectrogram of /atʃa/ in Quiet; 4 sine waves
(Frequency scale: 0-8820 Hz).

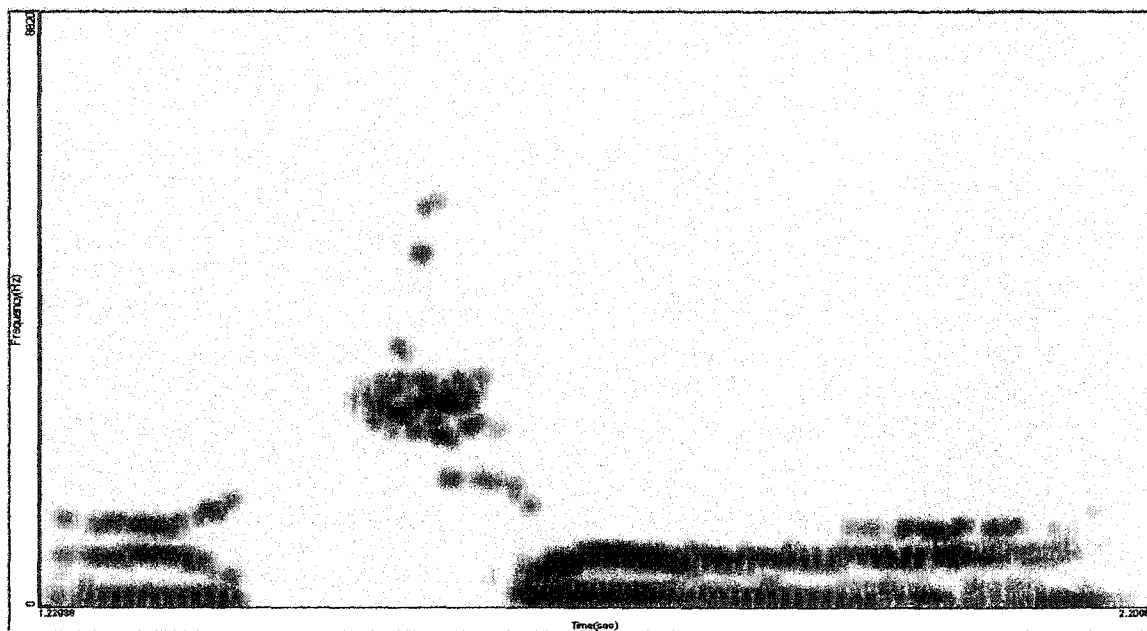


Figure 66. Wideband spectrogram of / atʃ a / in +10SNR; Unprocessed
(Frequency scale: 0-8820 Hz).

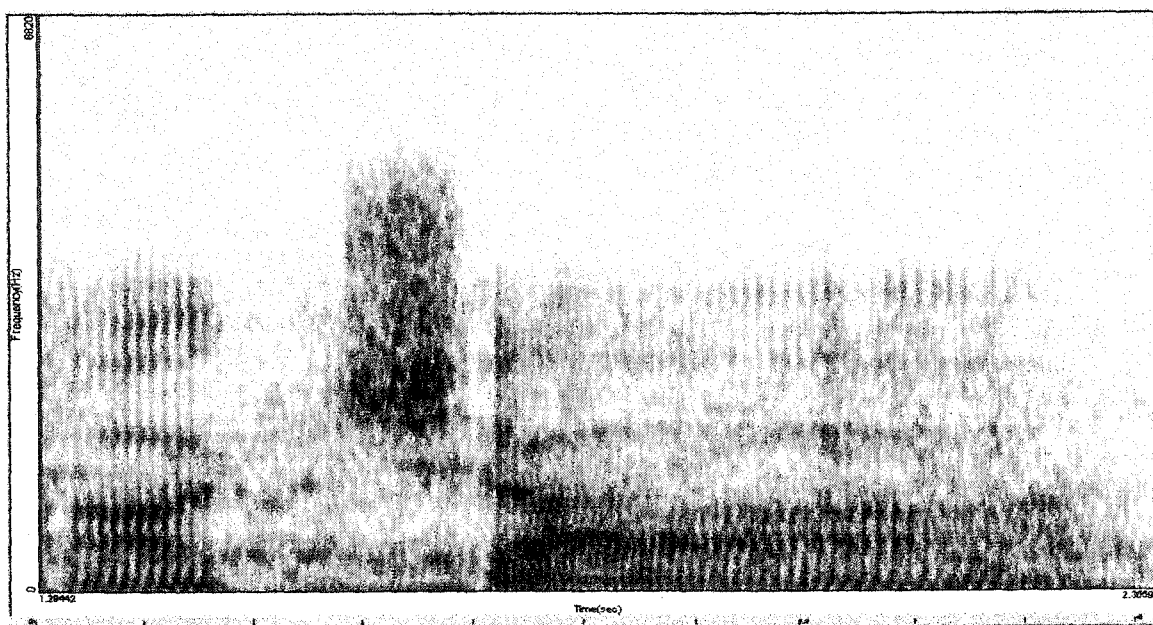


Figure 67. Wideband spectrogram of / atʃ a / in +10SNR:16 sine waves
(Frequency scale: 0-8820 Hz).



Figure 68. Wideband spectrogram of / at ζa / in +10SNR; 12 sine waves
(Frequency scale: 0-8820 Hz).



Figure 69. Wideband spectrogram of / at ζa / in +10SNR; 8 sine waves
(Frequency scale: 0-8820 Hz).



Figure 70. Wideband spectrogram of / atʃ a / in +10SNR; 4 sine waves
(Frequency scale: 0-8820 Hz).

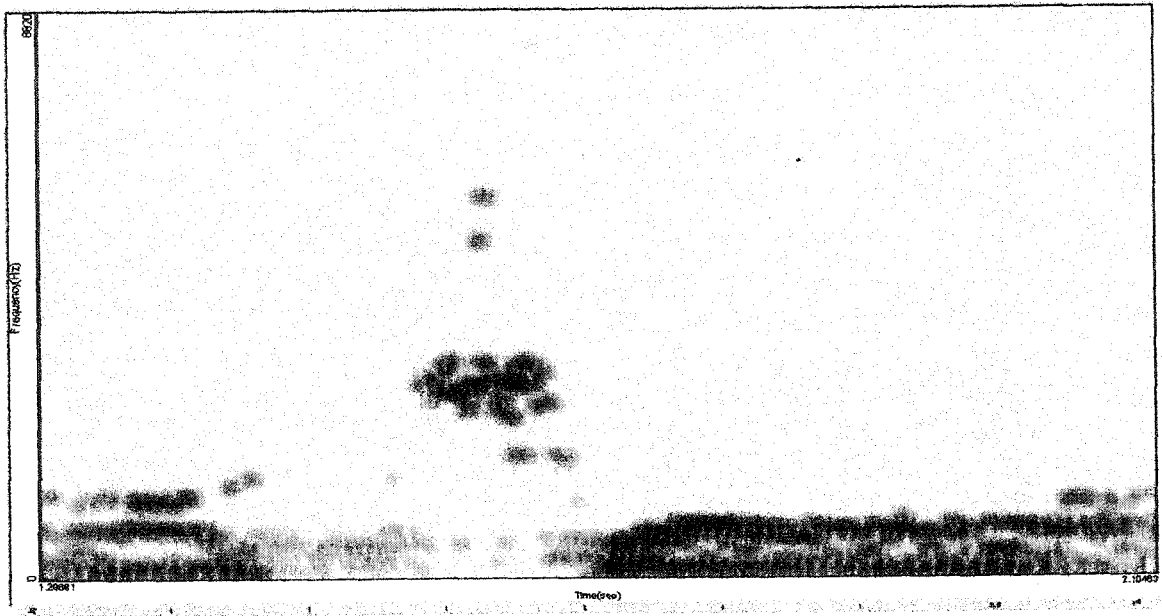


Figure 71. Wideband spectrogram of / atʃ a / in 0SNR; Unprocessed
(Frequency scale: 0-11020 Hz).



Figure 72. Wideband spectrogram of / atʃ a / in 0SNR; 16 sine waves
(Frequency scale: 0-11020 Hz).



Figure 73. Wideband spectrogram of / at ζ_a / in 0SNR; 12 sine waves
(Frequency scale: 0-11020 Hz).

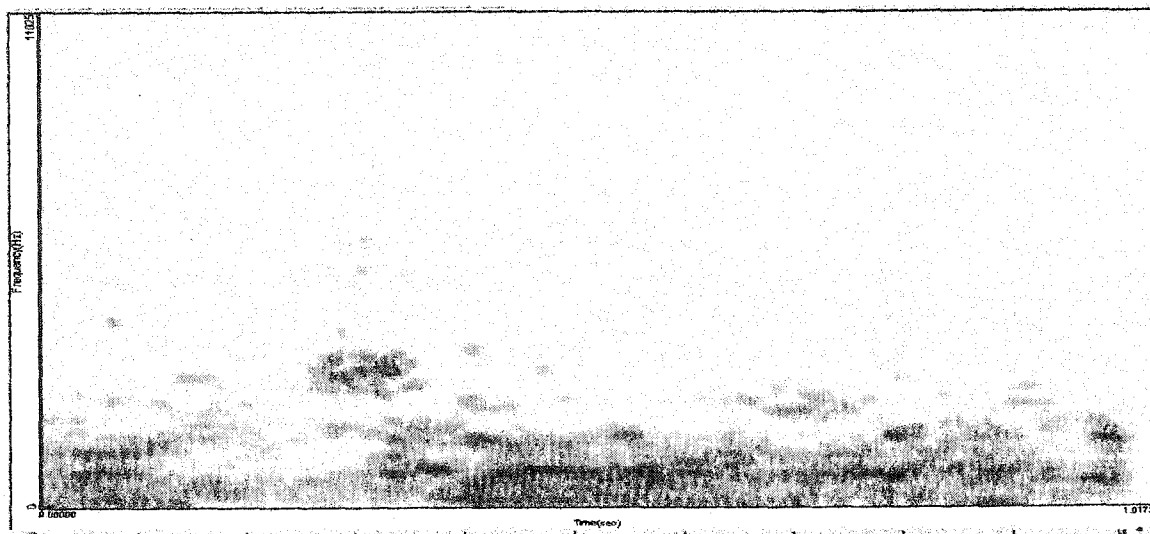
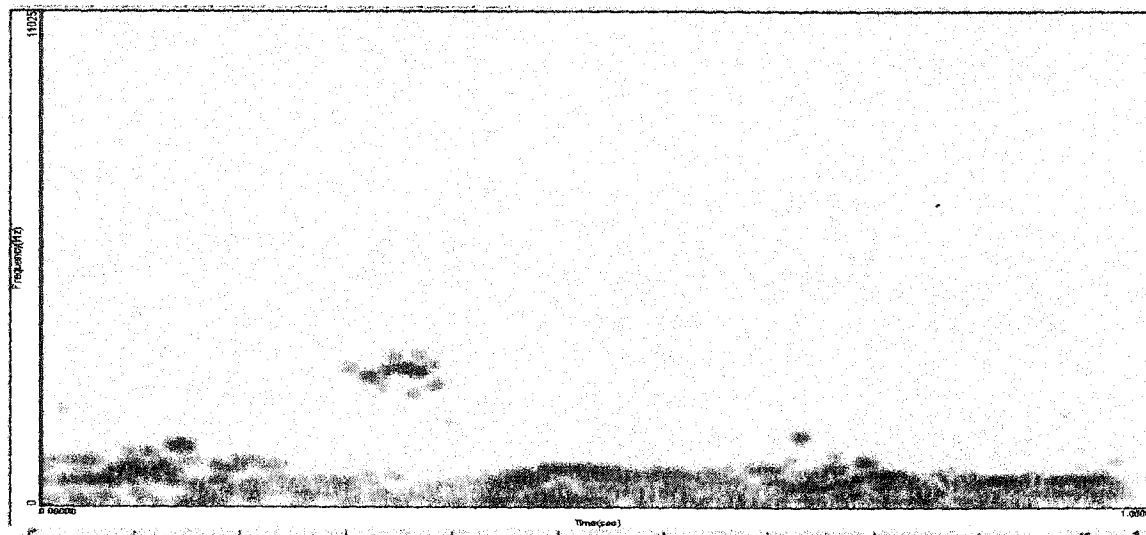


Figure 74. Wideband spectrogram of / at ζ_a / in 0SNR; 8 sine waves
(Frequency scale: 0-11020 Hz).



Figure 75. Wideband spectrogram of /atʃa/ in 0SNR; 4 sine waves
(Frequency scale: 0-11020 Hz).



Syllable /ana/

Figure 76. Wideband spectrogram of / ana / in Quiet; Unprocessed
(Frequency scale: 0-8820 Hz).

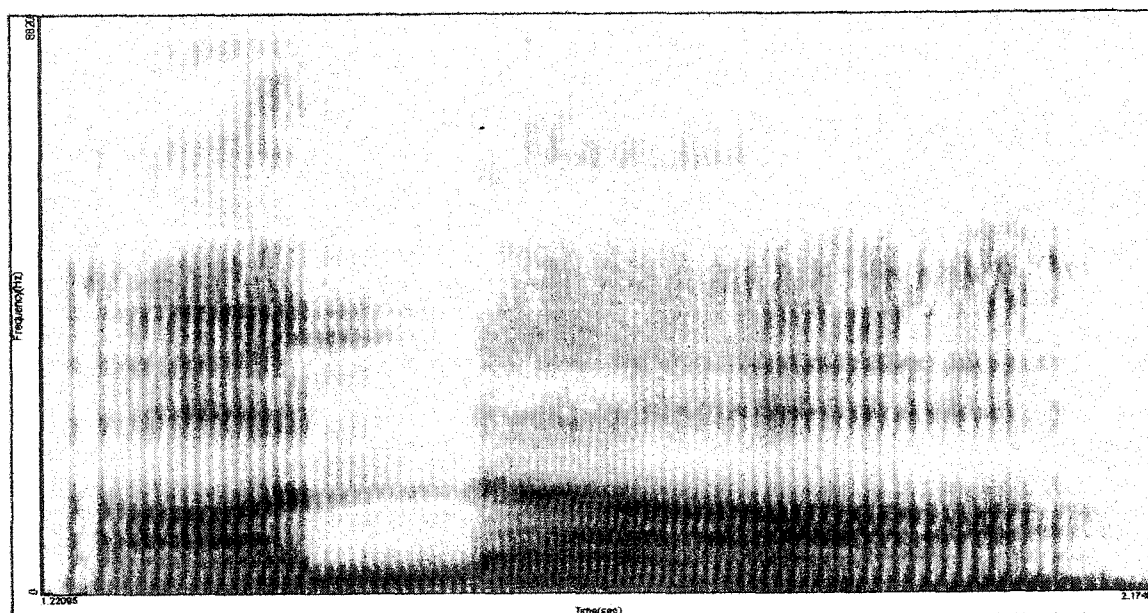


Figure 77. Wideband spectrogram of / ana / in Quiet; 16 sine waves
(Frequency scale: 0-8820 Hz).

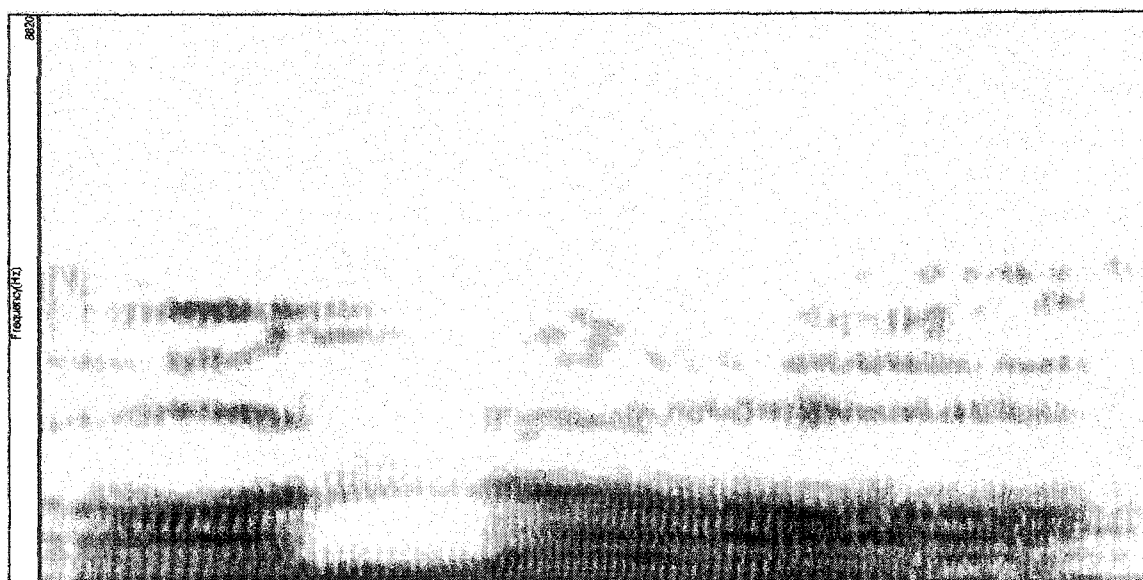


Figure 78. Wideband spectrogram of / ana / in Quiet; 12 sine waves
(Frequency scale: 0-8820 Hz).

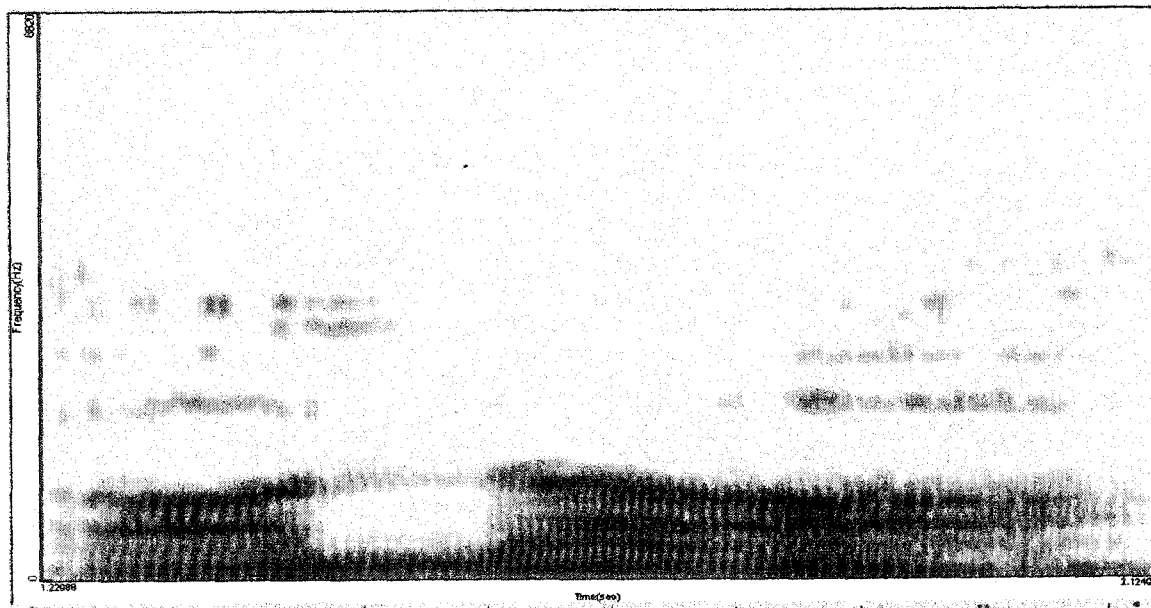


Figure 79. Wideband spectrogram of / ana / in Quiet; 8 sine waves
(Frequency scale: 0-8820 Hz).



Figure 80. Wideband spectrogram of / ana / in Quiet; 4 sine waves
(Frequency scale: 0-8820 Hz).

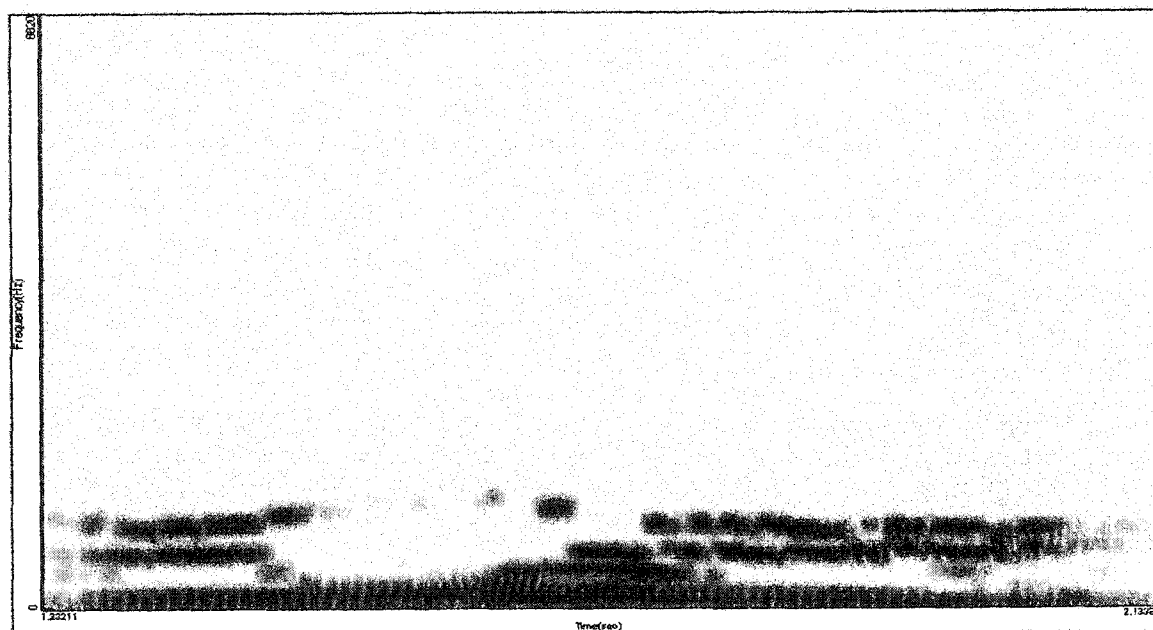


Figure 81. Wideband spectrogram of / ana / in +10SNR; Unprocessed
(Frequency scale: 0-8820 Hz).

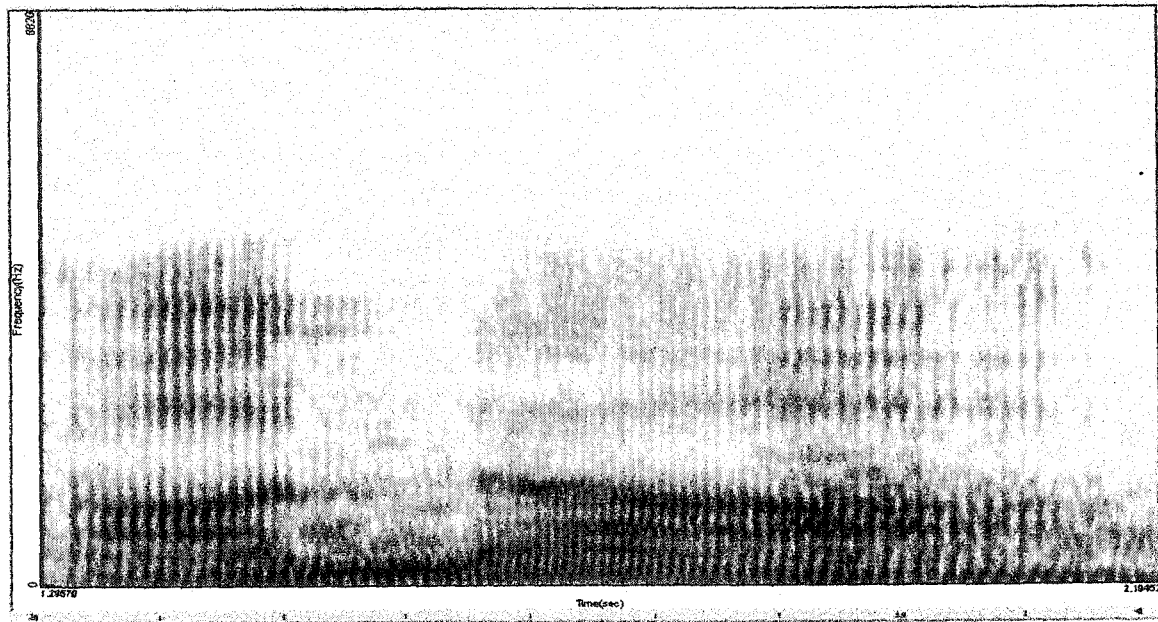


Figure 82. Wideband spectrogram of / ana / in +10SNR; 16 sine waves
(Frequency scale: 0-8820 Hz).



Figure 83. Wideband spectrogram of / ana / in +10SNR; 12 sine waves
(Frequency scale: 0-8820 Hz).

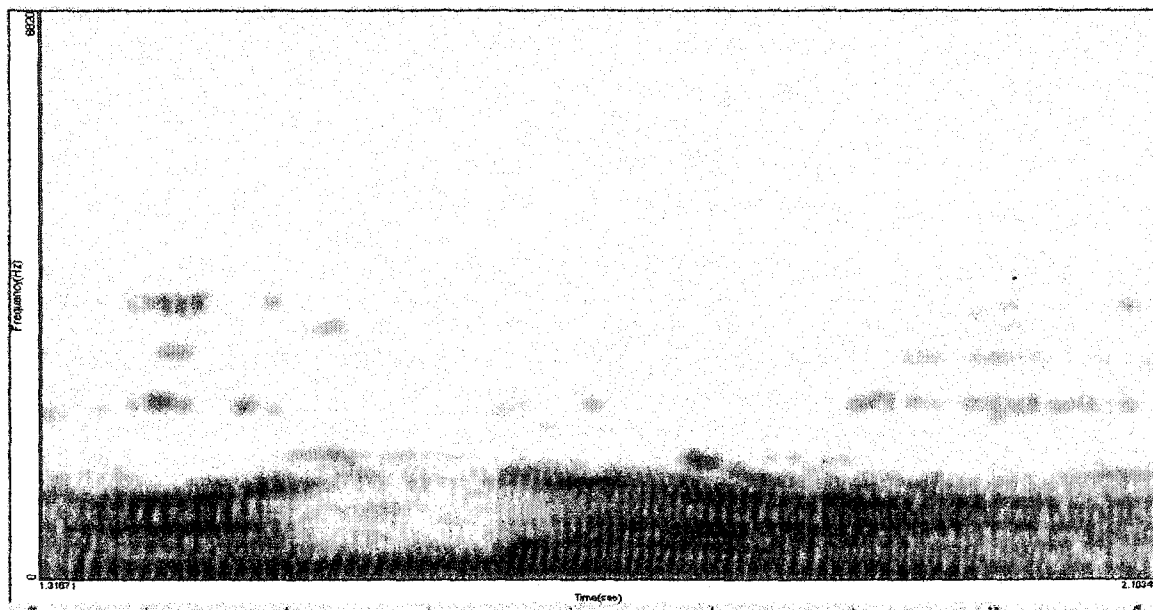


Figure 84. Wideband spectrogram of / ana / in +10SNR; 8 sine waves
(Frequency scale: 0-8820 Hz).

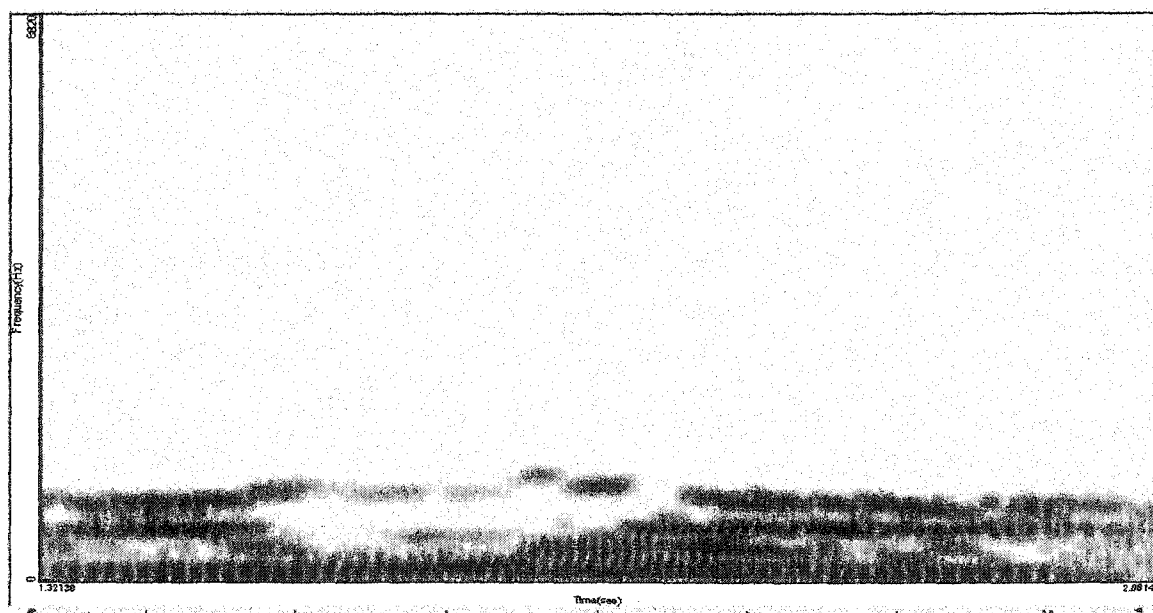


Figure 85. Wideband spectrogram of / ana / in +10SNR; 4 sine waves
(Frequency scale: 0-8820 Hz).

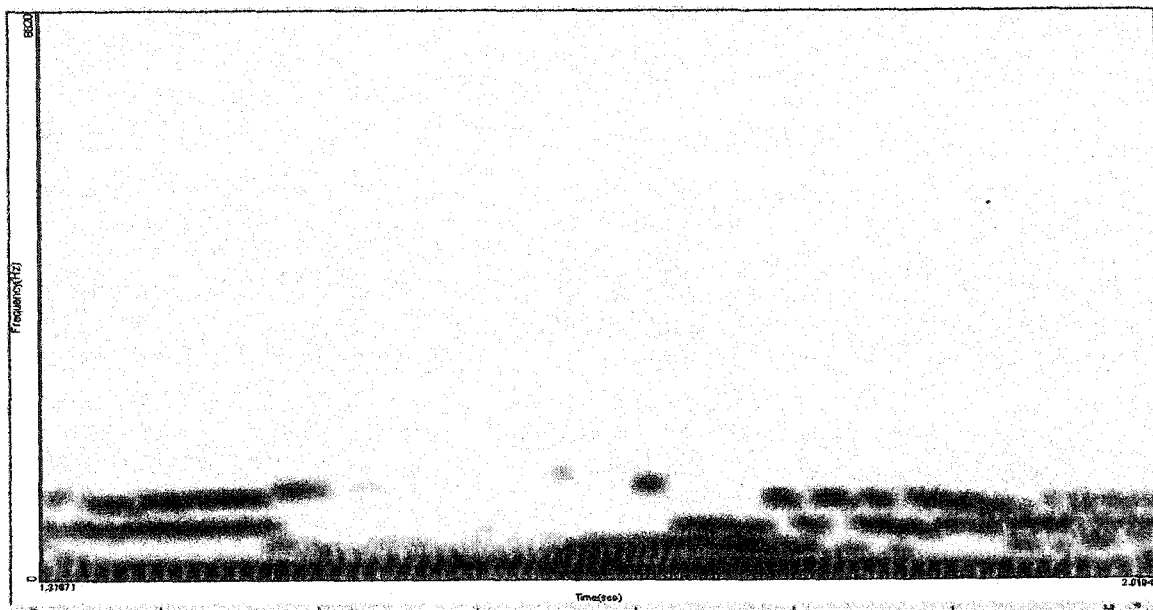


Figure 86. Wideband spectrogram of / ana / in 0SNR; Unprocessed
(Frequency scale: 0-11025 Hz).

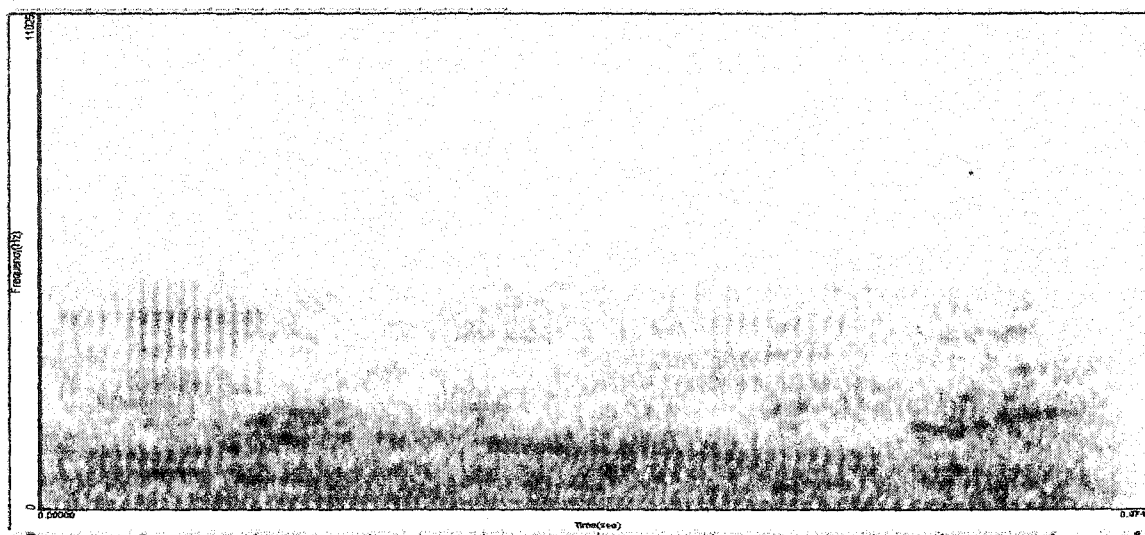


Figure 87. Wideband spectrogram of / ana / in 0SNR; 16 sine waves
(Frequency scale: 0-11025 Hz)



Figure 88. Wideband Wideband spectrogram of / ana / in 0SNR; 12 sine waves
(Frequency scale: 0-11025 Hz)

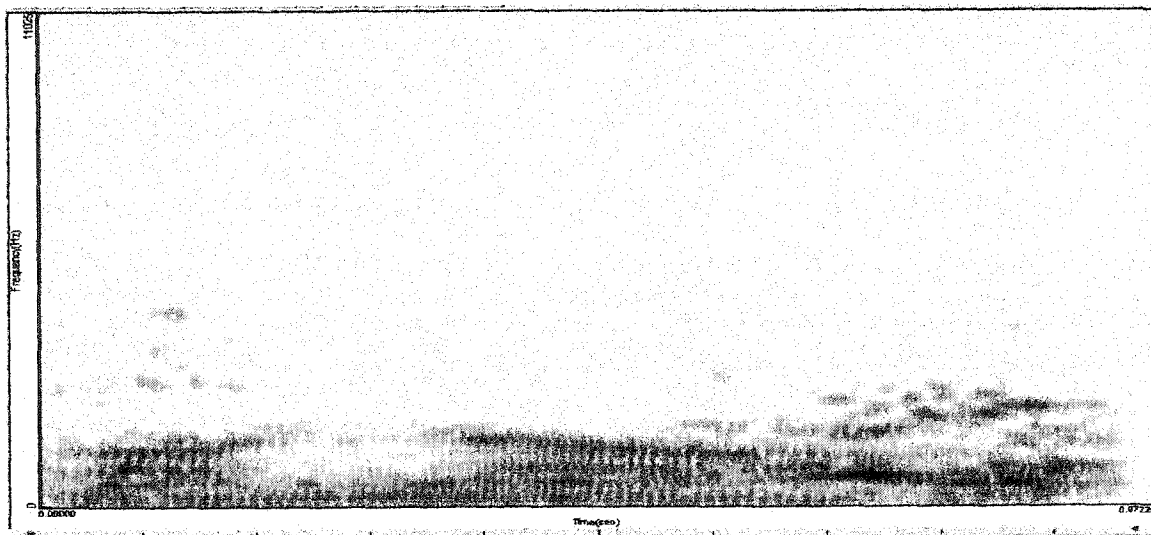


Figure 89. Wideband Wideband spectrogram of / ana / in 0SNR; 8 sine waves
(Frequency scale: 0-11025 Hz)

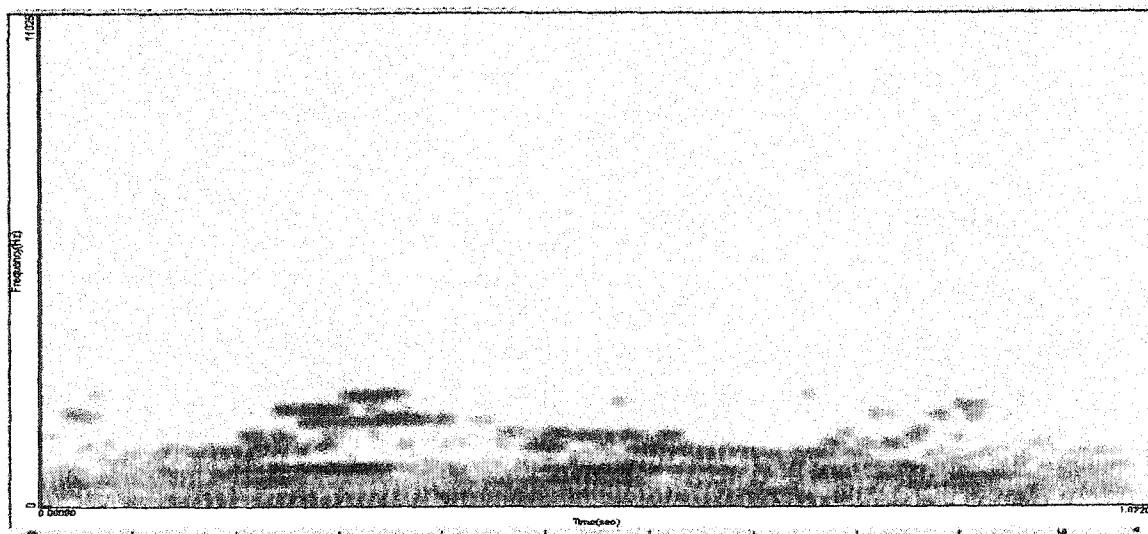
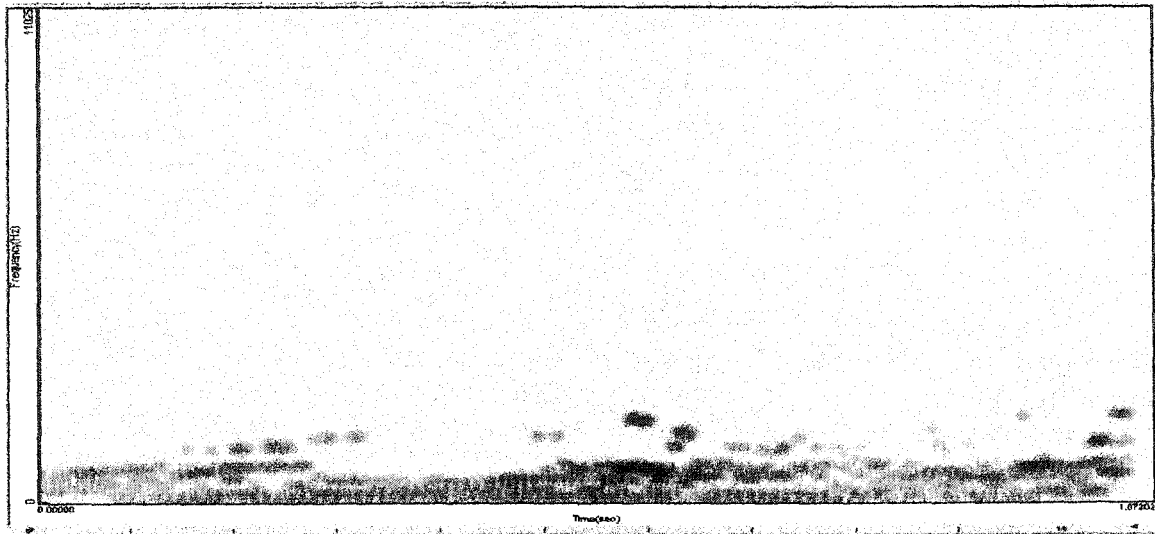


Figure 90. Wideband spectrogram of / ana / in 0SNR; 4 sine waves
(Frequency scale: 0-11025 Hz)



Syllable /ara/

Figure 91. Wideband spectrogram of / ara / in Quiet; Unprocessed
(Frequency scale: 0-8820 Hz)

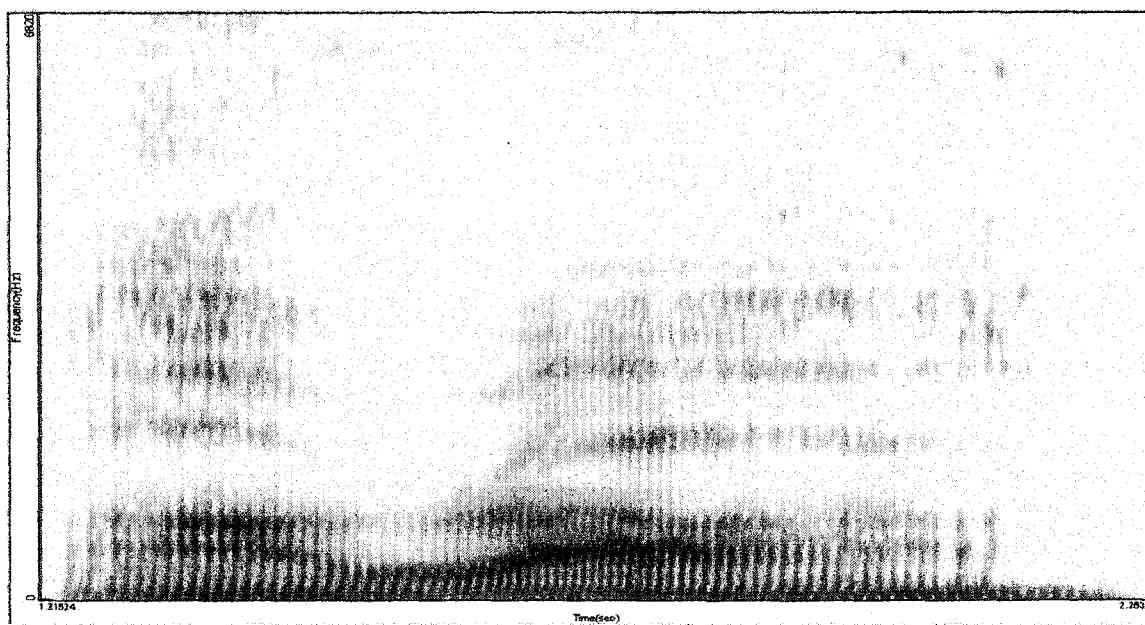


Figure 92. Wideband spectrogram of / ara / in Quiet; 16 sine waves
(Frequency scale: 0-8820 Hz)

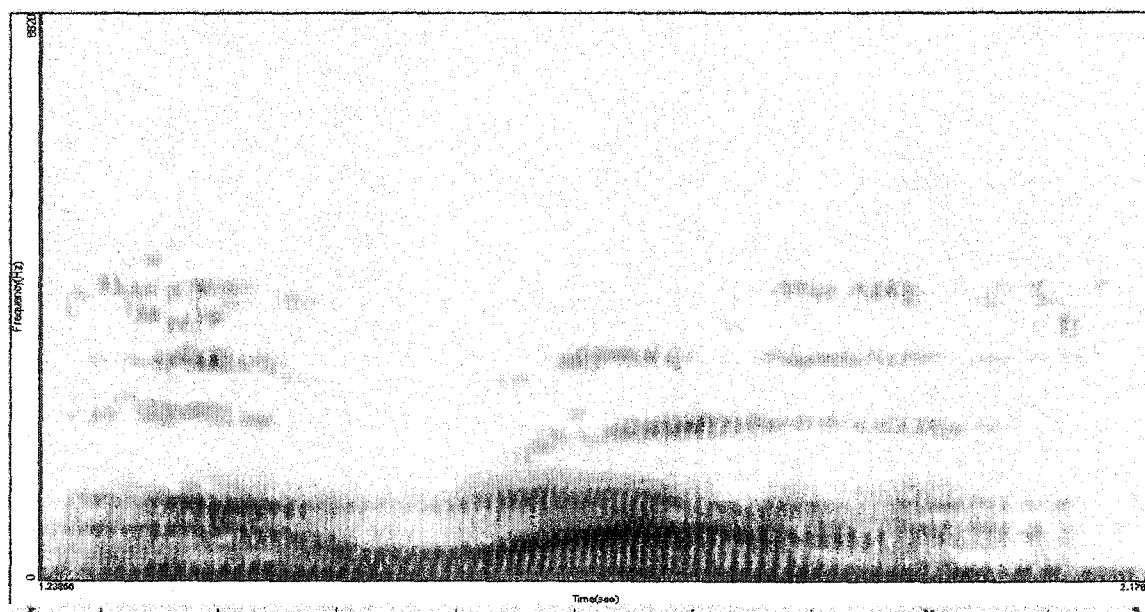


Figure 93. Wideband spectrogram of / ara / in Quiet; 12 sine waves
(Frequency scale: 0-8820 Hz)

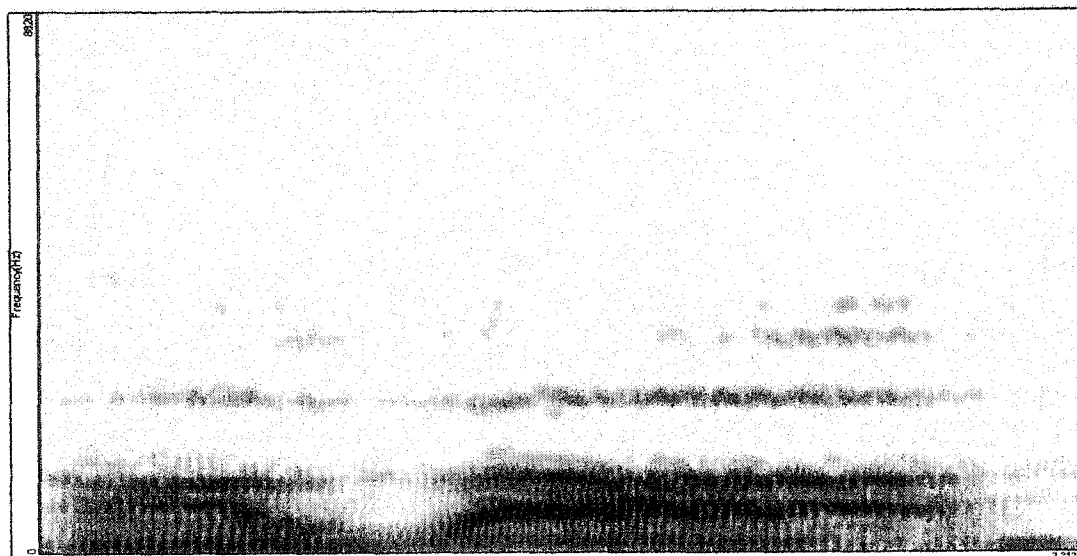


Figure 94. Wideband spectrogram of / ara / in Quiet; 8 sine waves
(Frequency scale: 0-8820 Hz)

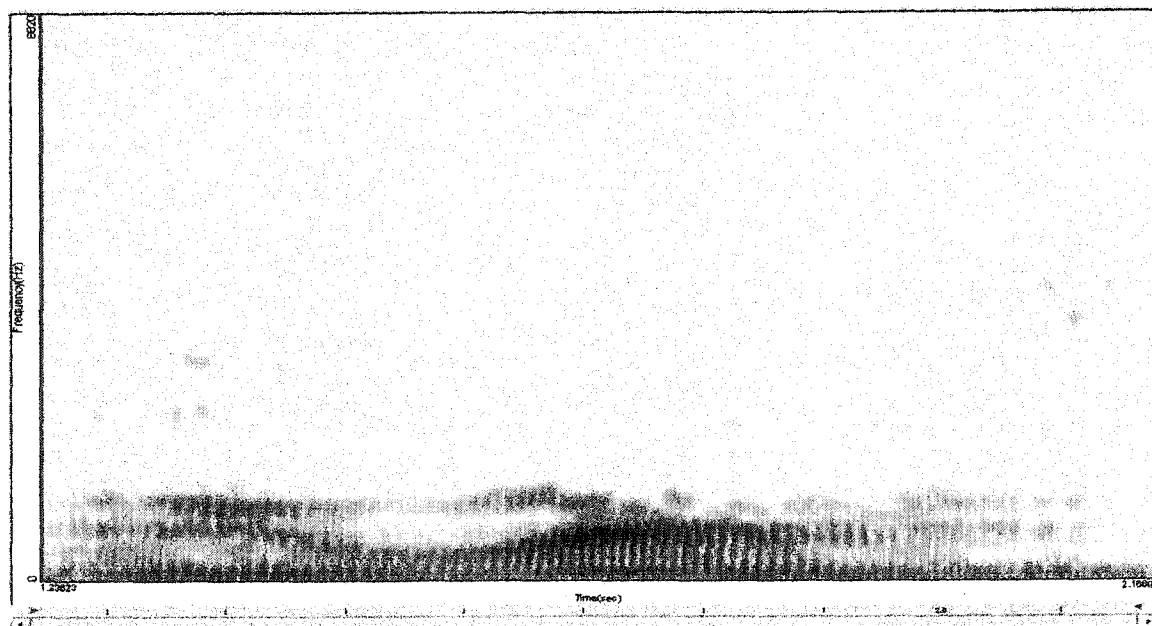


Figure 95. Wideband Wideband spectrogram of / ara / in Quiet; 4 sine waves
(Frequency scale: 0-8820 Hz)

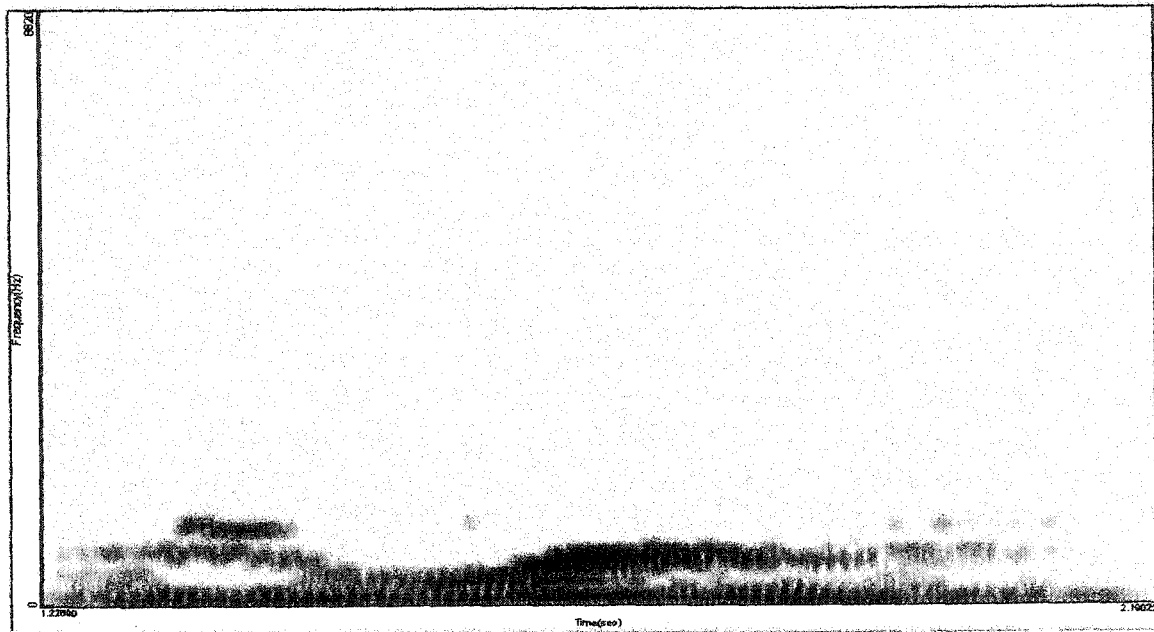


Figure 96. Wideband Wideband spectrogram of / ara / in +10SNR; Unprocessed
(Frequency scale: 0-8820 Hz)

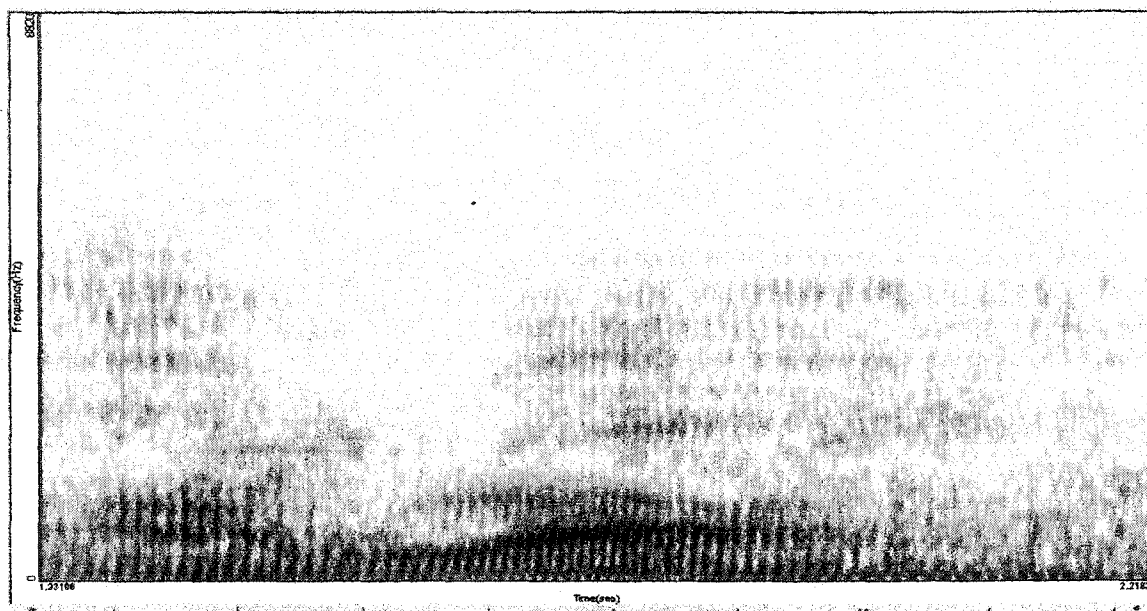


Figure 97. Wideband spectrogram of / ara / in +10SNR; 16 sine waves
(Frequency scale: 0-8820 Hz)

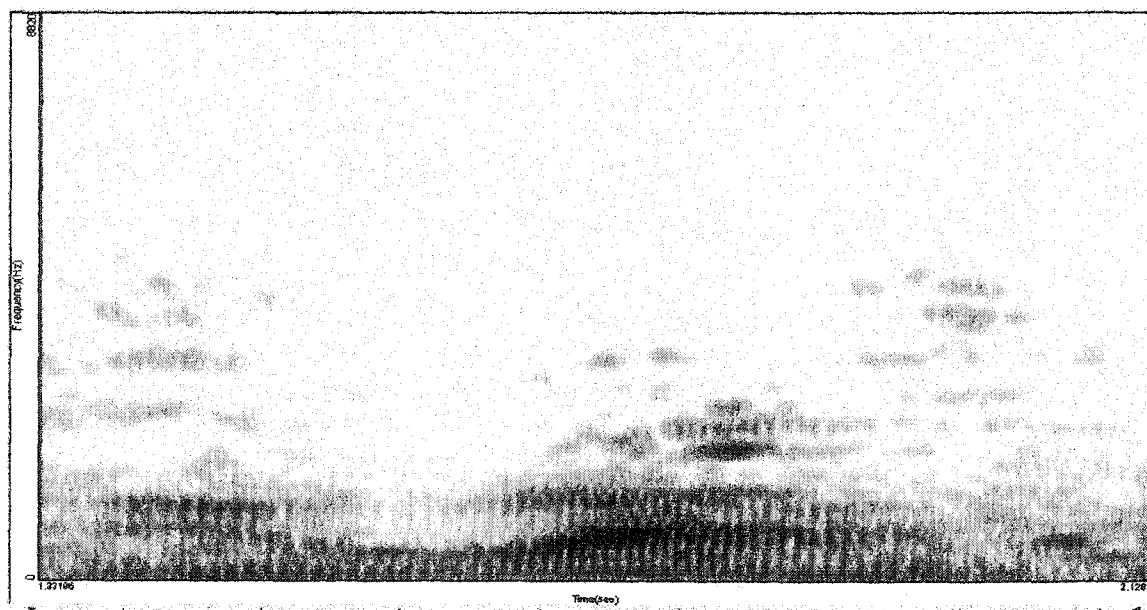


Figure 98. Wideband spectrogram of / ara / in +10SNR; 12 sine waves
(Frequency scale: 0-8820 Hz)

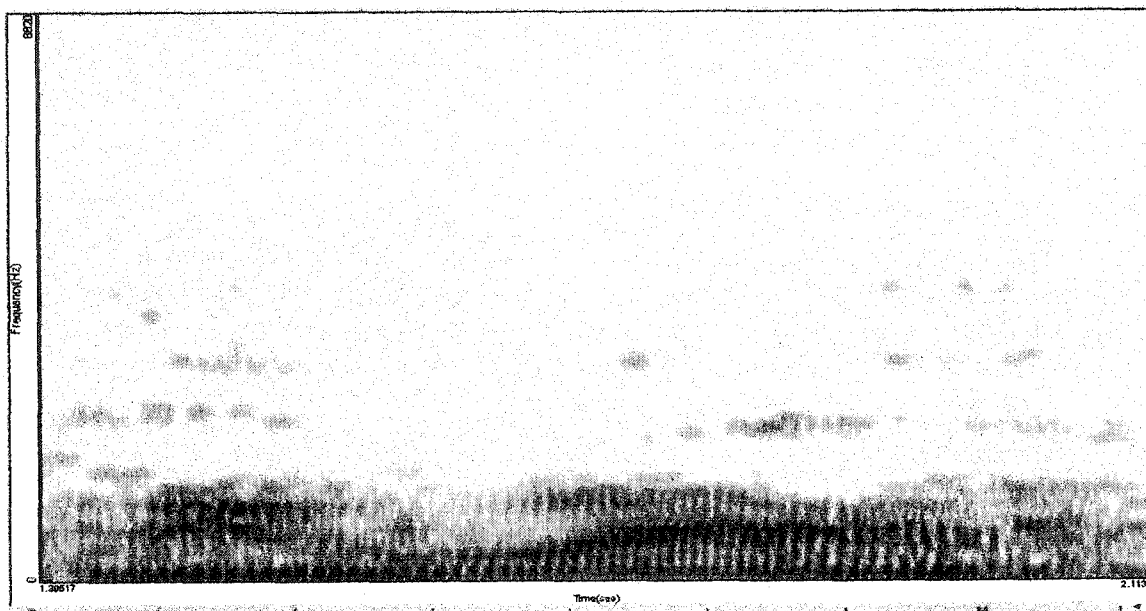


Figure 99. Wideband spectrogram of / ara / in +10SNR; 8 sine waves
(Frequency scale: 0-8820 Hz)

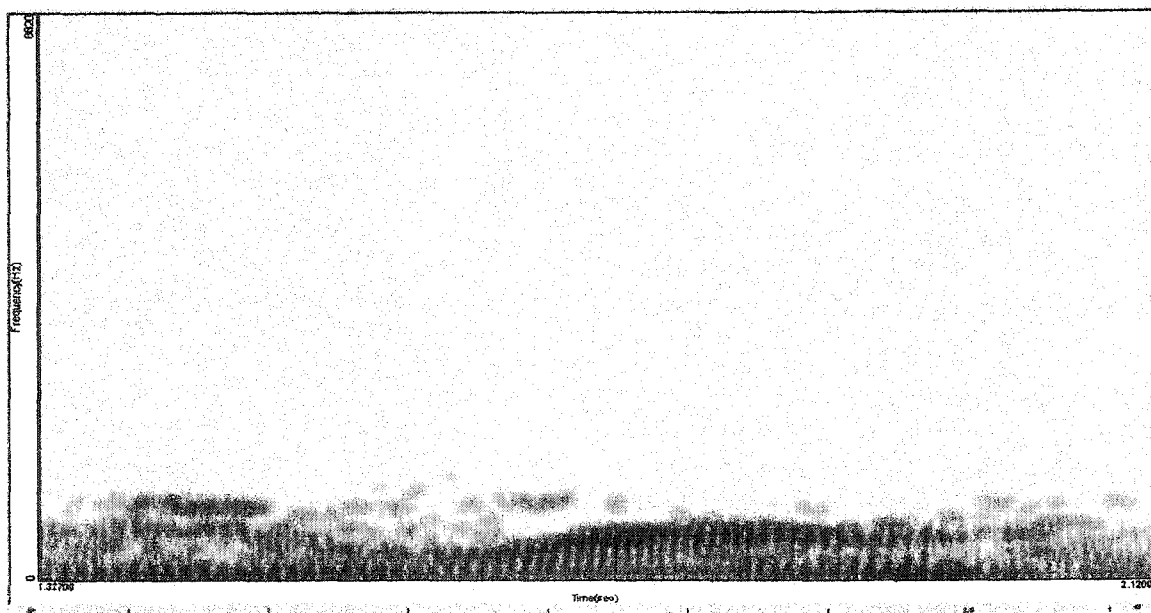


Figure 100. Wideband spectrogram of / ara / in +10SNR; 4 sine waves
(Frequency scale: 0-8820 Hz)

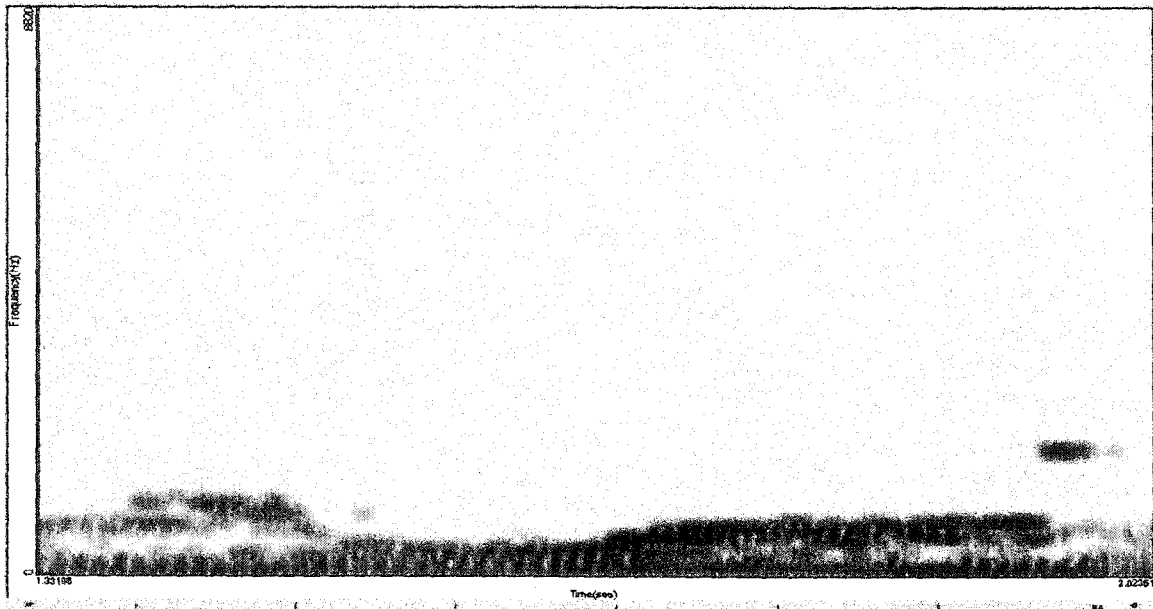


Figure 101. Wideband spectrogram of / ara / in 0SNR; Unprocessed
(Frequency scale: 0-11025 Hz)

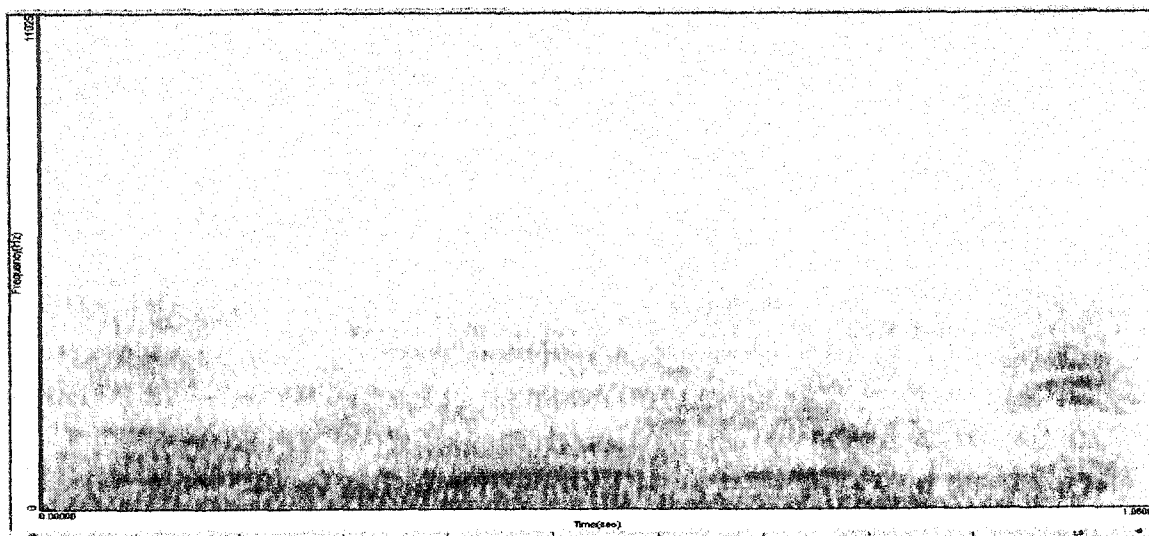


Figure 102. Wideband spectrogram of / ara / in 0SNR; 16 sine waves
(Frequency scale: 0-11025 Hz)



Figure 103. Wideband spectrogram of / ara / in 0SNR; 12 sine waves
(Frequency scale: 0-11025 Hz)

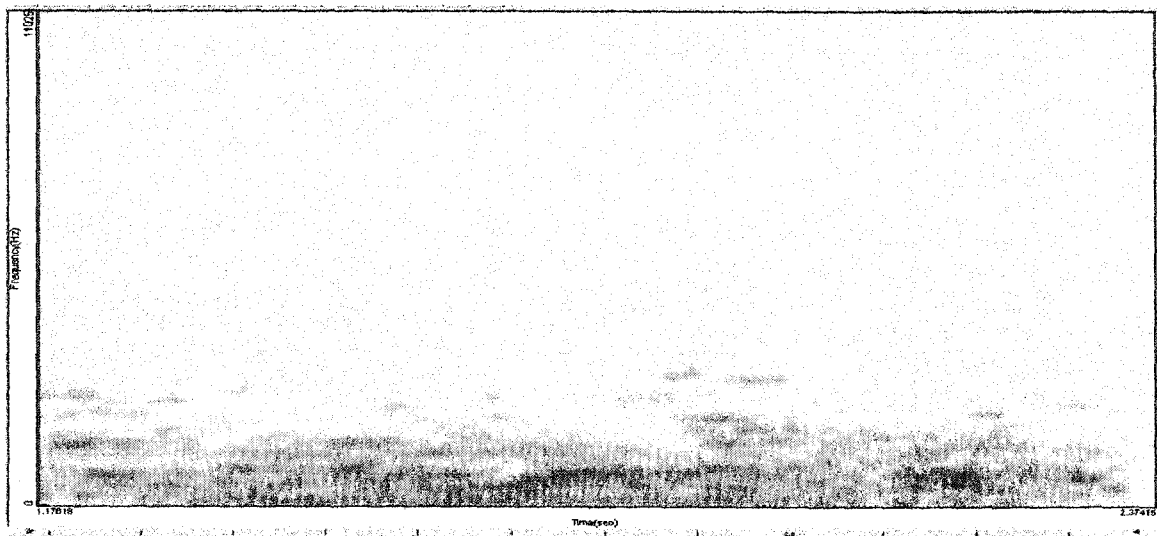


Figure 104. Wideband spectrogram of / ara / in 0SNR; 8 sine waves
(Frequency scale: 0-11025 Hz)

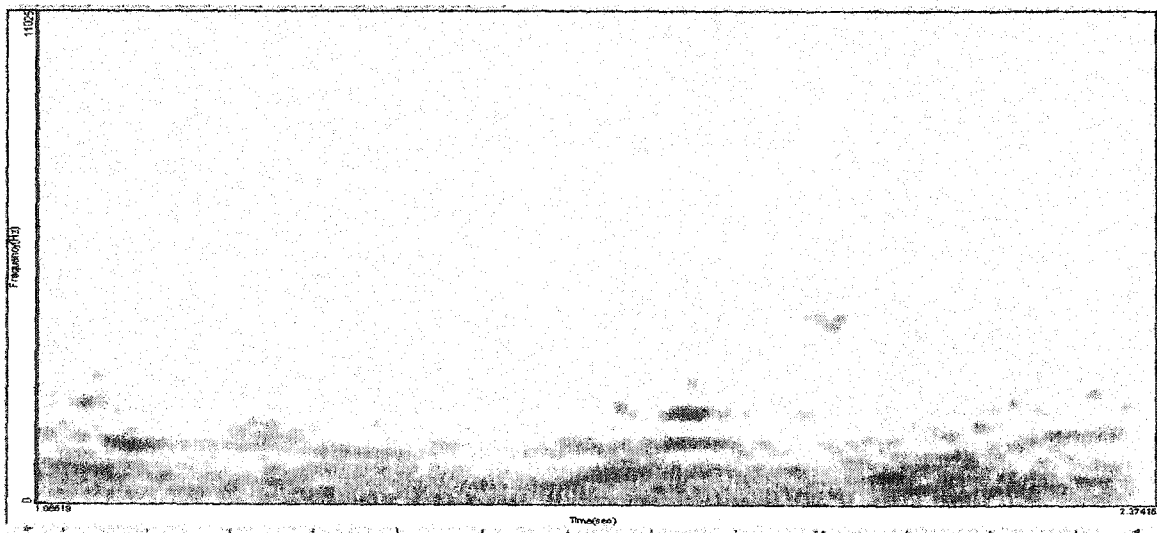
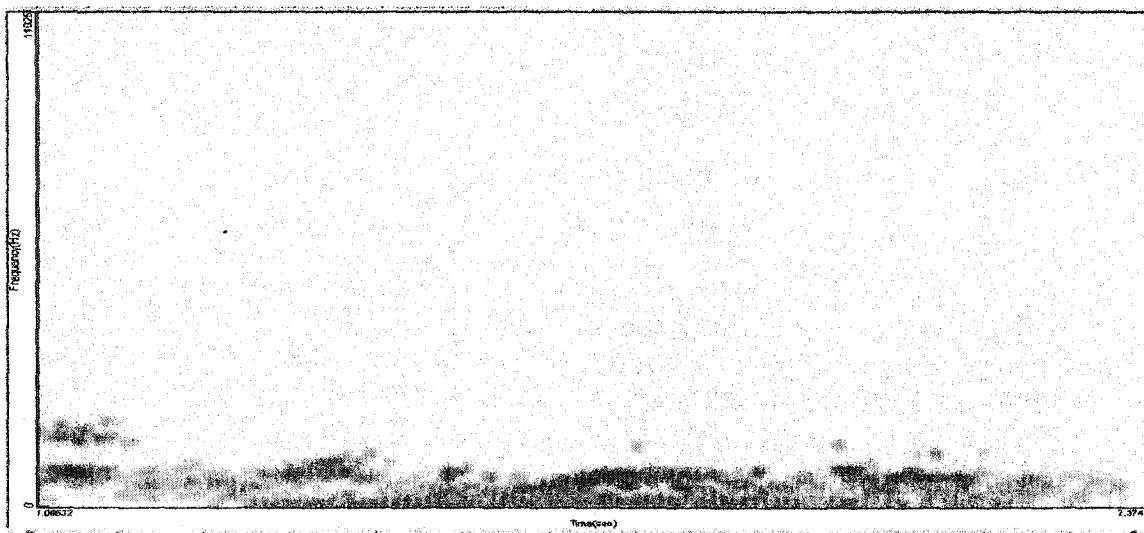


Figure 105. Wideband spectrogram of / ara / in 0SNR; 4 sine waves
(Frequency scale: 0-11025 Hz)



Syllable (/aja/)

Figure 106. Wideband spectrogram of / aja / in Quiet; Unprocessed
(Frequency scale: 0-8820 Hz)

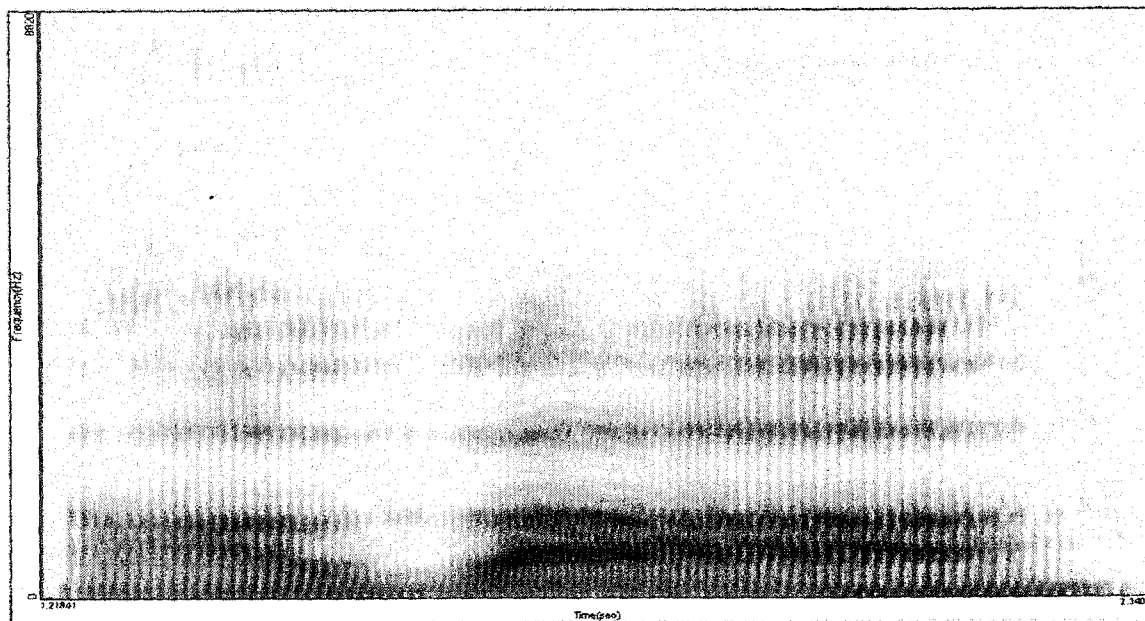


Figure 107. Wideband spectrogram of / aja / in Quiet; 16 sine waves
(Frequency scale: 0-8820 Hz)

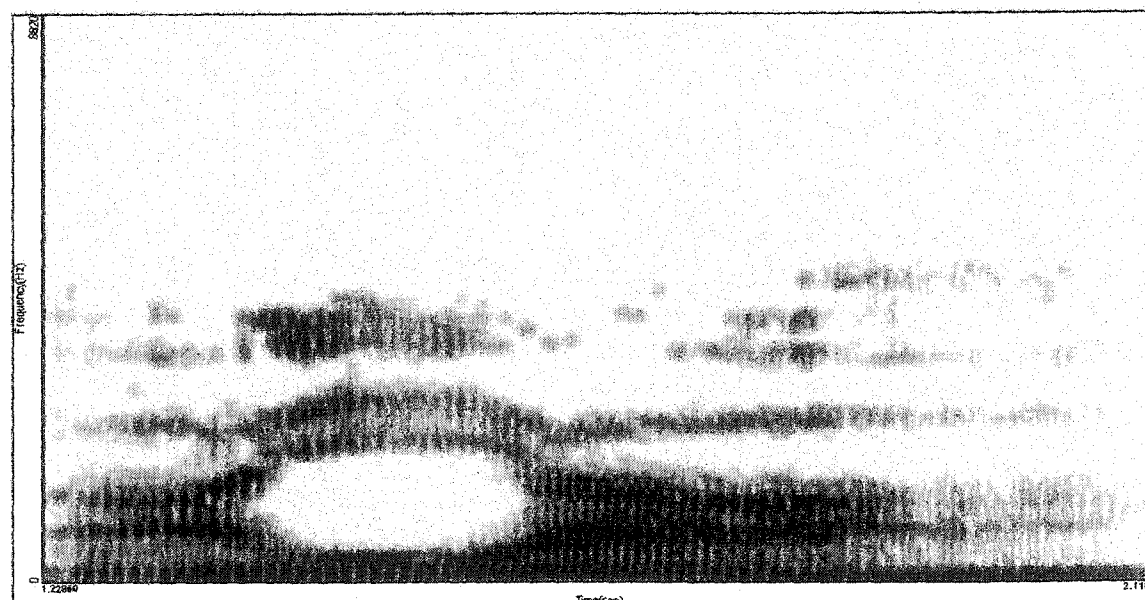


Figure 108. Wideband spectrogram of / aja / in Quiet; 12 sine waves
(Frequency scale: 0-8820 Hz)



Figure 109. Wideband spectrogram of / aja / in Quiet; 8 sine waves
(Frequency scale: 0-8820 Hz)

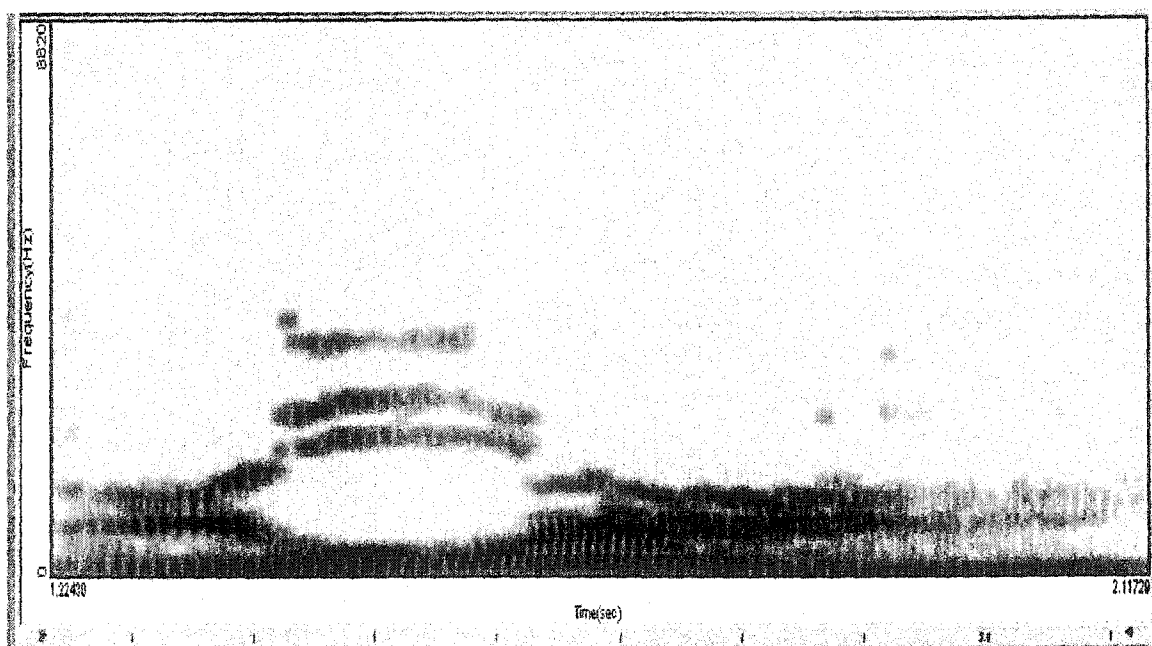


Figure 110. Wideband spectrogram of / aja / in Quiet; 4 sine waves
(Frequency scale: 0-8820 Hz)



Figure 111. Wideband spectrogram of / aja / in +10SNR; Unprocessed
(Frequency scale: 0-8820 Hz)

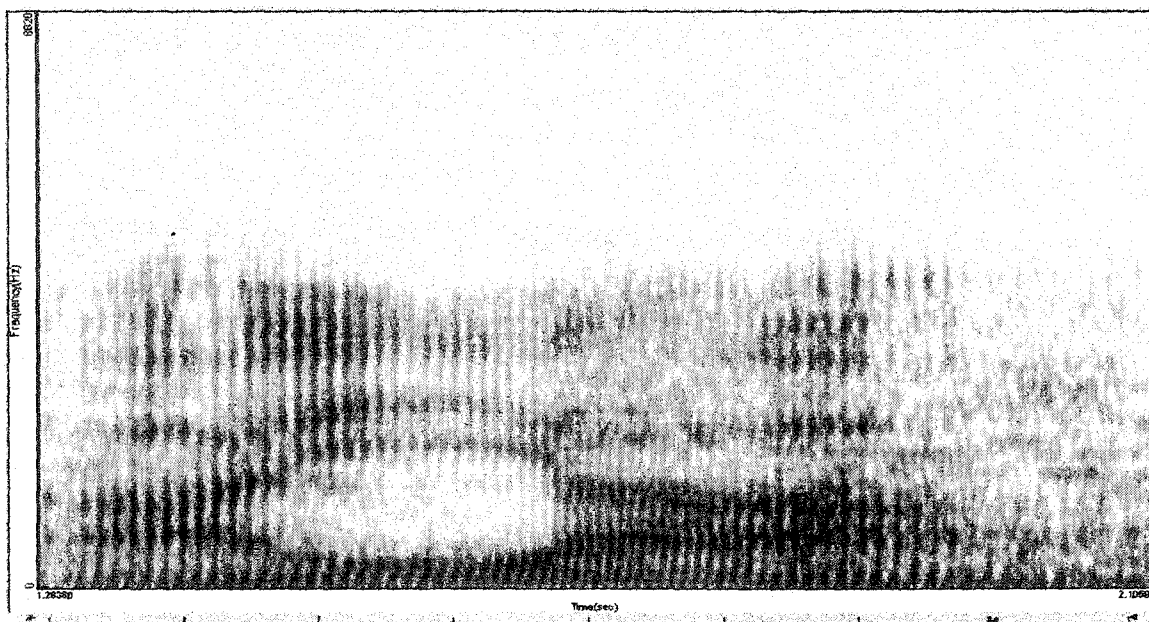


Figure 112. Wideband spectrogram of / aja / in +10SNR; 16 sine waves
(Frequency scale: 0-8820 Hz)

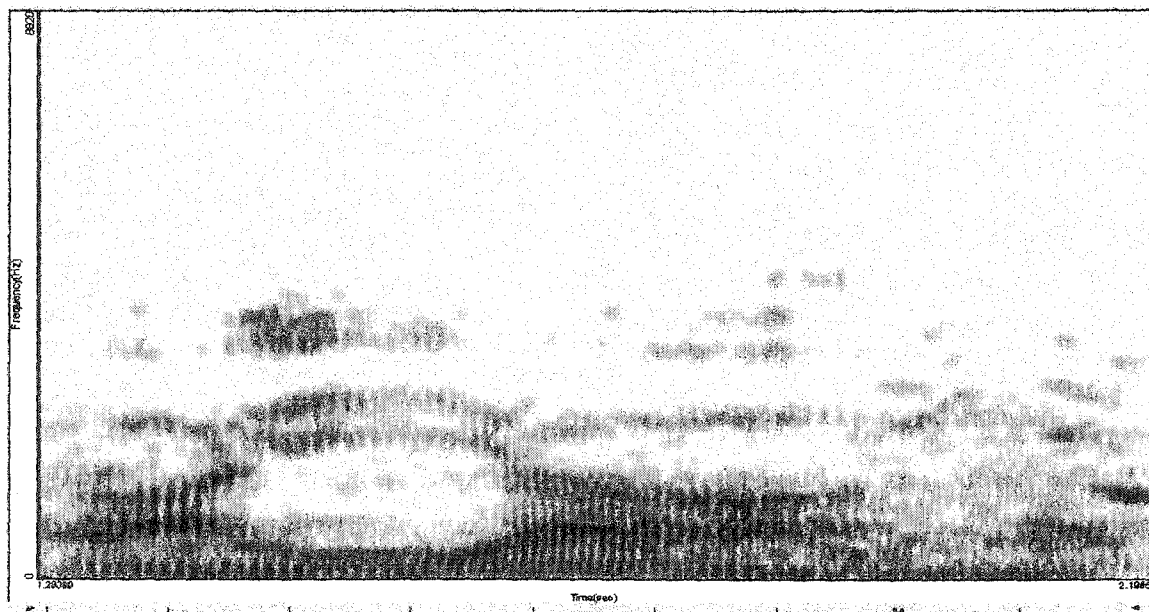


Figure 113. Wideband spectrogram of / aja / in +10SNR; 12 sine waves
(Frequency scale: 0-8820 Hz)

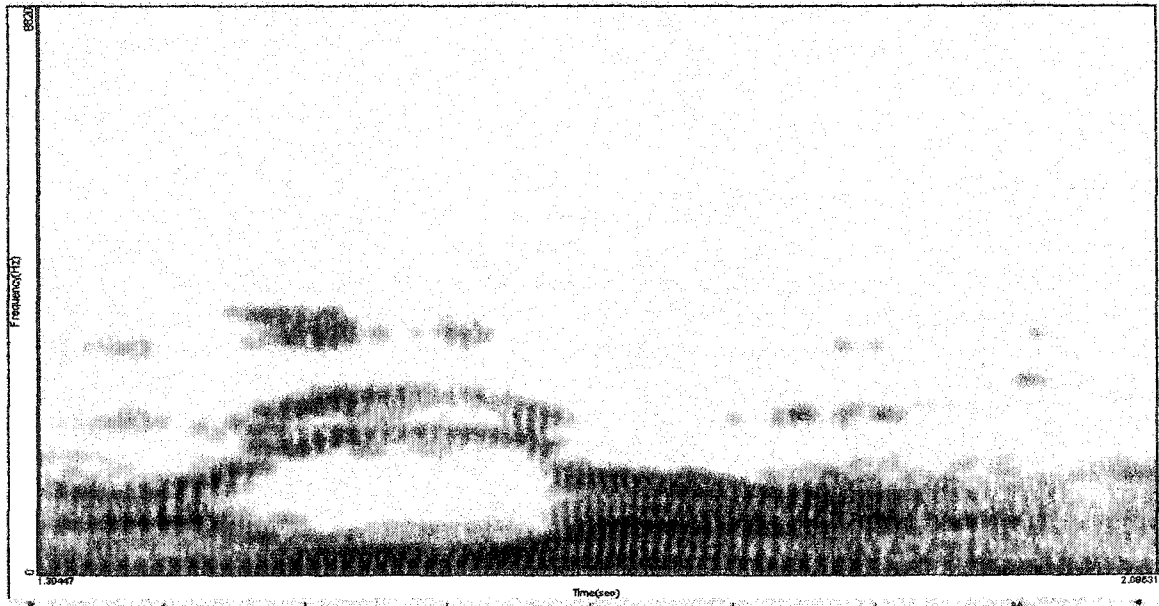


Figure 114. Wideband spectrogram of / aja / in +10SNR; 8 sine waves
(Frequency scale: 0-8820 Hz)

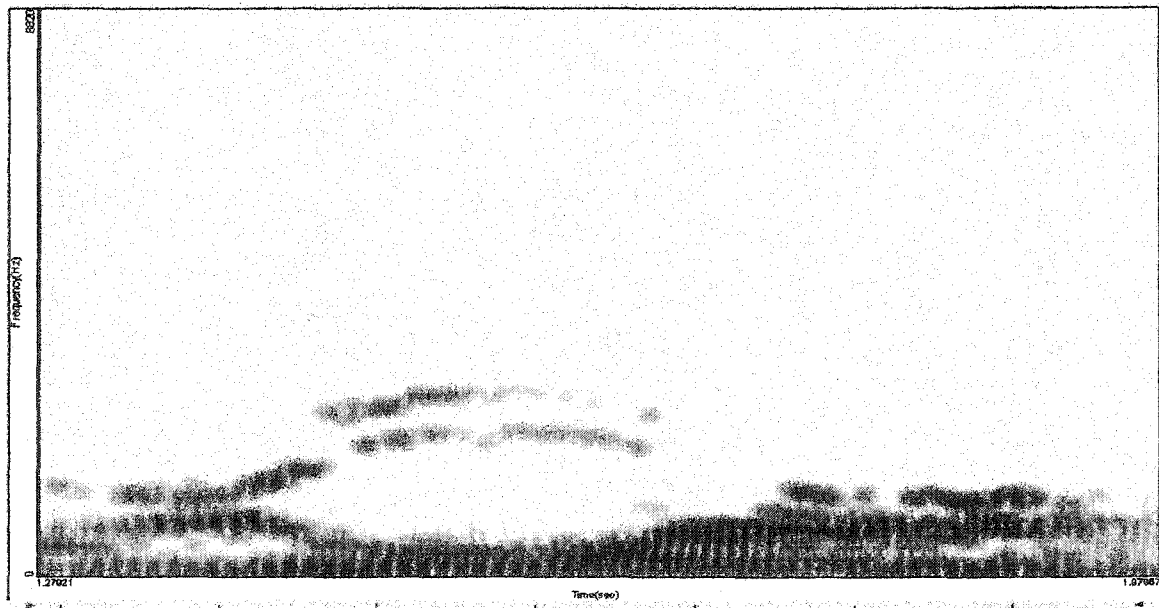


Figure 115. Wideband spectrogram of / aja / in +10SNR; 4 sine waves
(Frequency scale: 0-8820 Hz)



Figure 116. Wideband spectrogram of / aja / in 0SNR; Unprocessed
(Frequency scale: 0-11025 Hz)

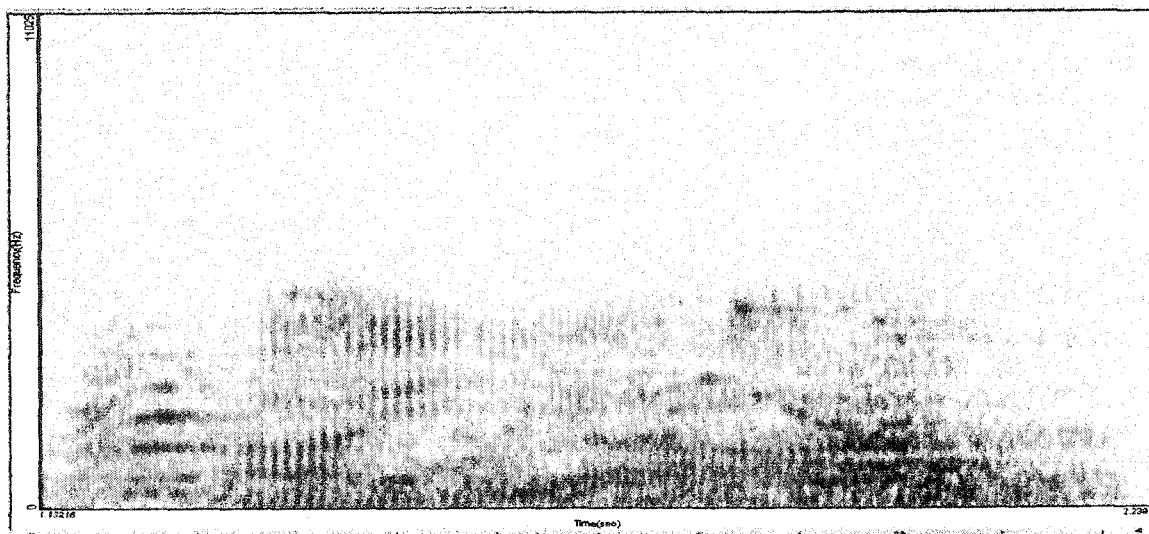


Figure 117. Wideband spectrogram of / aja / in 0SNR; 16 sine waves
(Frequency scale: 0-11025 Hz)



Figure 118. Wideband spectrogram of / aja / in 0SNR; 12 sine waves
(Frequency scale: 0-11025 Hz)



Figure 119. Wideband spectrogram of / aja / in 0SNR; 8 sine waves
(Frequency scale: 0-11025 Hz)

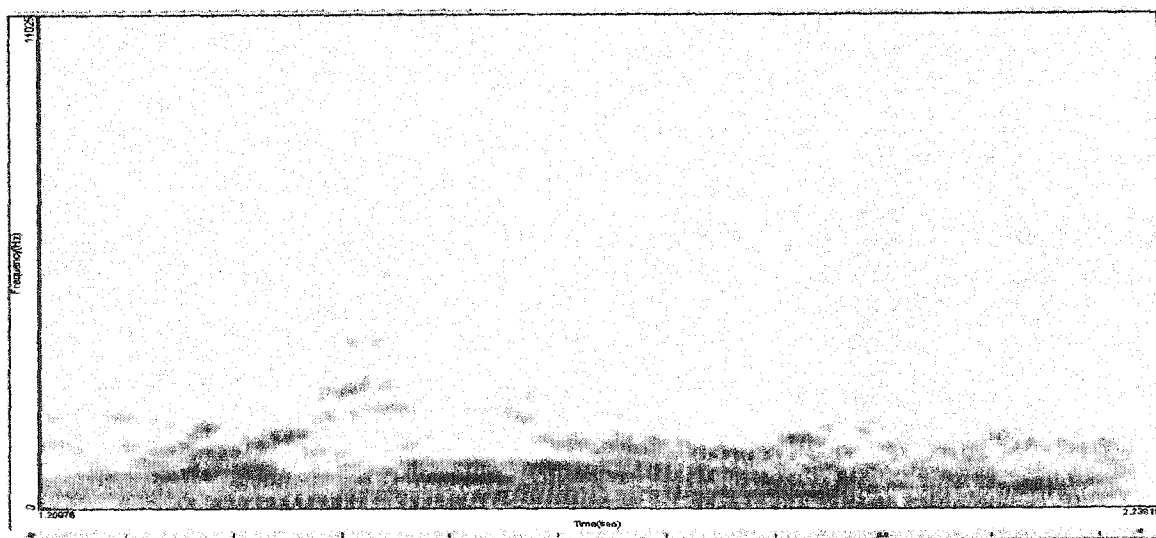
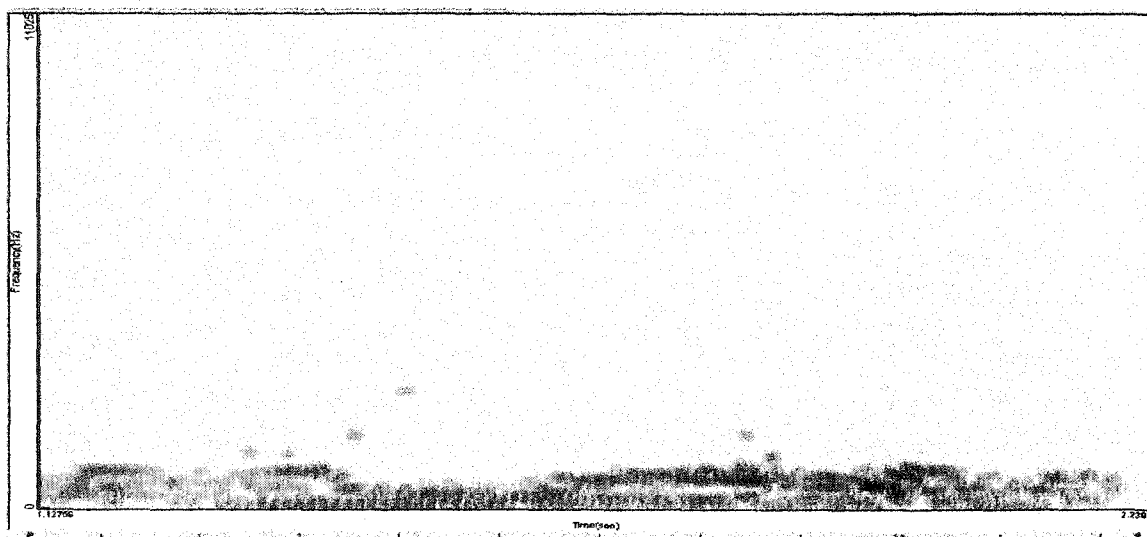


Figure 120. Wideband spectrogram of / aja / in 0SNR; 4 sine waves
(Frequency scale: 0-11025 Hz)



APPENDIX 2

Confusion Matrices for Normal Hearing AdultsTable 1Confusion Matrix for Normal hearing adults for unprocessed stimuli in Quiet

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	35		1																	
t		36																		
k			36																	
b				36																
d					36															
g						36														
f							36													
θ							1	34	1											
s									36											
ʃ										36										
v											33								1	
z								2				36								
tʃ													36							
dʒ						4								32						
m															36					
n																36				
l																	36			
r																		36		
w																			36	
j																				36

Table 2

Confusion Matrix for Normal hearing adults for 16 sine waves in Quiet

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	34		2																	
t		36																		
k			36																	
b	1			33							1							1		
d					36															
g				1		35														
f							32	4												
θ								30	6											
s									36											
ʃ		1								35										
v							1				33								1	1
z								1				35								
tʃ													36							
dʒ						1								35						
m															36					
n																36				
l																	36			
r																		36		
w																			36	
j																				36

Table 3

Confusion Matrix for Normal hearing adults for 12 sine waves in Quiet

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	35		1																	
t		36																		
k			36																	
b	3			32							1									
d					36															
g	1					35														
f							33	3												
θ								28	8											
s									36											
ʃ								1		35										
v							1	1			30								3	1
z								2				34								
tʃ													36							
dʒ						1								35						
m															36					
n																36				
l																	36			
r																		36		
w																			36	
j																				36

Table 4

Confusion Matrix for Normal hearing adults for 8 sine waves in Quiet

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	32		4																	
t		36																		
k			36																	
b	2			32							1							1		
d					36															
g						36														
f							31	5												
θ								24	11			1								
s									36											
ʃ										36										
v						1					30							1	3	1
z								2				34								
tʃ													36							
dʒ						1								35						
m															36					
n																36				
l																	36			
r						1												35		
w																			36	
j															1					35

Table 5

Confusion Matrix for Normal hearing adults for 4 sine waves in Quiet

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	24		12																	
t		34											2							
k			34										2							
b	3			25	1	2					4								1	
d					34	1								1						
g	1			1		27								7						
f					1		30	4	1											
θ		1						11	21			3								
s									36											
ʃ										36										
v								2			18	3						3	7	3
z												36								
tʃ													35	1						
dʒ						2								34						
m															28	5		1	2	
n															4	24	4		3	1
l						1										3	28	1	3	
r								1			3						2	19	11	
w																			36	
j																				36

Table 6

Confusion Matrix for Normal hearing adults for unprocessed stimuli at +10 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	30		6																	
t		36																		
k			36																	
b				32		1					3									
d					36															
g					1	35														
f							31	5												
θ								34	2											
s									36											
ʃ										36										
v								3			33									
z												36								
tʃ			1										35							
dʒ														36						
m															32	2	1	1		
n																34	2			
l																1	35			
r																	1	34	1	
w																		4	32	
j															1					35

Table 7

Confusion Matrix for Normal hearing adults for 16 sine waves at +10 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	31		5																	
t		36																		
k	2		34																	
b				28		3					5									
d					31	5														
g					1	35														
f	6		2	2			10	16												
θ								33	3											
s									36											
ʃ										35			1							
v				4		4		1			26								1	
z								2				34								
tʃ													36							
dʒ						1								35						
m															36					
n											1				2	31	2			
l																2	31	3		
r																		34	2	
w											2							12	22	
j																	4			32

Table 8

Confusion Matrix for Normal hearing adults for 12 sine waves at +10 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	31		4				1													
t		35				1														
k	5		31																	
b	1			29		2					4									
d					33	3														
g						36														
f	7		3	5			9	11			1									
θ			1	1		2		30	2											
s									36											
ʃ										35			1							
v				4		5		1			24								2	
z								2				34								
tʃ													36							
dʒ						1								35						
m															34			2		
n															5	29	2			
l																3	31	2		
r											1							34	1	
w											1							17	17	
j															1		4	1		30

Table 9

Confusion Matrix for Normal hearing adults for 8 sine waves at +10 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	28		5				3													
t		35											1							
k	6		29														1			
b	1			27		4		1			3									
d					28	8														
g				1	1	33								1						
f	10		2	2			11	10			1									
θ		2	2			1	1	30												
s								1	35											
ʃ										35			1							
v				4		3				1	22							3	2	1
z								2				33								1
tʃ													36							
dʒ						1							1	34						
m											1				33	1		1		
n															8	23	1	3	1	
l															4	3	25	4		
r								2			1							30	3	
w											1							15	19	
j								1							1	2	1	1	1	29

Table 10

Confusion Matrix for Normal hearing adults for 4 sine waves at +10 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	25		6	1			3	1												
t		34											2							
k	8		26	2																
b				16	5	6	3				6									
d	2			5	17	7					2		1	1						1
g				1	3	23		1			3			4					1	
f	10			5		2	9	8			1		1							
θ	3	5	4	2	1		7	13										1		
s									36											
ʃ						1				32			3							
v				4		4	4				14							1	7	2
z					1	1		4			10	8			1		1	5	2	3
tʃ													36							
dʒ						2								34						
m											1				23	7	3		2	
n							1								7	16	9	1	2	
l											1				1	5	21	2	6	
r											2				1		2	14	17	
w											3					2	1	7	23	
j											1					3	2	6	8	16

Table 11
Confusion Matrix for Normal hearing adults for unprocessed speech at 0 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	24		10				1	1												
t		31											5							
k	4	1	27				1	3												
b	3		1	20	2	1	1	1			6								1	
d					29	5					1		1							
g			1		3	32														
f			1				24	8	3											
θ					1			30	3	2										
s									36											
ʃ							1		1	34										
v			1	2		1	1	4			24						2			1
z												36								
tʃ													36							
dʒ						1						1		34						
m															20	5	8	1	1	1
n																22	10		1	3
l											1						31	3	1	
r					2	2					8						1	21	2	
w											4					3		8	20	1
j															1	5	4			26

Table 12

Confusion Matrix for Normal hearing adults for 16 sine waves at 0 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	25		8				1	2												
t		31	4										1							
k	2	4	25			3	1	1												
b	1		1	23		1	1	1			8									
d	1				18	16					1									
g	2		1		1	28		1			1			1						1
f	14	1	7	1			7	6												
θ	3	4	2	3				17	5			1	1							
s									35			1								
ʃ		1	1							24			10							
v				13	3			2			16							2		
z								1				35								
tʃ		4	2							3			26			1				
dʒ					3	1						1		30						1
m						1					2				25		3	2	2	1
n															2	24	8	1		1
l						3					1				2	11	16	3		
r								1			3	1					1	25	5	
w						1	1				4				2	1	1	11	15	
j					1	1	1				1				2	2	6		1	22

Table 13

Confusion Matrix for Normal hearing adults for 12 sine waves at 0 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	17	1	6	1		1	3	7												
t		28	1										6	1						
k	12	1	19				1	3												
b	2	3		18		5		1			7									
d		1	3		25	6								1						
g			1	3	8	19	1							3		1				
f	8		9	5	3		4	7												
θ	4	1	1		1	2	4	23												
s									36											
ʃ								1		27			7	1						
v				2	2	4		2			19	1						6		
z								3				32					1			
tʃ		1	3		1								31							
dʒ			1			4					1			28					2	
m								1			1				22	6	2	3	1	
n				1				1							4	16	3	8	3	
l					1						1			1	2	12	17	1		1
r											1							30	5	
w								1							2	2	1	15	13	2
j											1			1	6	2	2	5	3	16

Table 14

Confusion Matrix for Normal hearing adults for 8 sine waves at 0 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	18	1	8				4	5												
t		35	1																	
k	4	3	23					2		2										
b	5			13	4	2	1	3		1	6				1					
d		1		2	17	11					4			1						
g	1			3	4	21								2				1	1	3
f	12	3	3	3		1	6	8												
θ	4	2	3				9	17			1									
s									35			1								
ʃ	1		2					1		23			9							
v	1	3		5	3	3		4			11	2	1					2		1
z				1	2	5		5			2	12		3				3	1	2
tʃ			2					1					32	1						
dʒ		1			3	4								27						1
m															22	7		6		1
n											1				3	14	5	6	4	3
l			1	1	1	2					2		1		2	6	14	4	2	
r				1		1		1			5			2	1		1	21	2	1
w							1	1			3				4		4	8	14	1
j				2	1			1							2	6	5	4	3	12

Table 15

Confusion Matrix for Normal hearing adults for 4 sine waves at 0 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	14	8	3				5	3					3							
t	4	17	3		1		6	2					3							
k	9	6	9				2	3					7							
b	3		1	9	9	5		4			5									
d	1	1	1	2	14	9		3			1	3		1						
g	2				11	14	2	2	1		1		1	2						
f	14	1	3	1		1	8	5	3											
θ	6	3	3		3		6	9	4			1						1		
s	2	3	2				1	1	19			7	1							
ʃ	2	4	1			1		1		6			15	6						
v			1	3	5	9	1	3		1	7	3		2					1	
z				1	7	4		3			1	12		1		1	1		1	4
tʃ	1	5							2	3			22	3						
dʒ		1			1	1		1					2	29						1
m											1			1	11	15	5	2	1	
n															6	14	5	3	3	5
l						3		2			3				5	4	7	3	8	1
r					1	1		1			3	1				1	7	5	10	6
w					1						6				1	2	2	7	13	4
j				1				2				1		1	1	2	4	4	8	9

APPENDIX 3

Confusion matrices for the hearing-impaired subjects

Table 1

Confusion Matrix For Hearing-impaired listeners: Unprocessed Stimuli in Quiet

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	85	7	26						1			1								
t	1	95	18					1					5							
k	3	8	105		1							1	2							
b			1	94	6	10		1			4	2						1	1	
d			1	1	90	25					1	1		1						
g			1		5	107						3	2	1	1					
f	3						87	8	16	1		1	4							
θ	5	10	5	1			28	35	28	1	1	1	5							
s		1	1				3	3	95	3		8	4			1				1
ʃ										113			6				1			
v				7		1	3	3	2		79	21			1				3	
z			1		1		1	2	2		21	83					1	1	5	2
tʃ		6	1			1		1					109	2						
dʒ			1		5	26				1		1	2	84						
m											4	2			89	8	15	1		1
n						1	1	1			1				21	75	18	1	1	
l			1								6	1			5	3	101	1	1	1
r						1		2	2		21	7					9	58	16	4
w											11	4					6	25	74	
j		1				1		1		2	6	3		4	1		1	1	1	98

Table 2

Consonant confusion Matrix for hearing-impaired subjects, 16 sine waves quiet

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	78	14	24			1	1									1		1		
t	1	84	11			1		1		1		2	16	3						
k	5	7	99					1			1		5		1					1
b	1			76	5	14	4	2	1		12	2		1		1				1
d				1	78	28	1	1	1		1	1		6		1				1
g	2				10	98	1				3	2		4						
f	2				1		87	7	14	4		2	3							
θ	4	8	3	2	1		25	36	30		1	5	4	1						
s		3	1			1	4	6	95	2	3	3	2							
ʃ		2					1	1	1	109			5			1				
v			1	5		3		2	1		75	22	1	1				8	1	
z						2		5	2		24	78					1	1	3	4
tʃ		1				4			2	4			102	6						1
dʒ			1		5	27			1			1	1	84						
m											9	3			81	9	13	1	3	1
n										1	2				23	71	17	2	4	
l			1					1			8	3			3	2	95	2	2	3
r			1	1				2			26	6					12	52	19	1
w								1			12	3	1	1			10	23	67	2
j								1	1	2	4	4		5	1	1	2	3	2	94

Table 3

Consonant confusion Matrix for hearing-impaired subjects, 12 sine waves quiet

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	81	11	25										2	1						
t	2	76	12		2		1	2	1	1			20	3						
k	6	8	95			1		1					7	1		1				
b	2			71	11	9	1	3	1		16	4	1						1	
d			2		75	30					3	2	1	6						1
g		1	2	1	4	95	1			1	7	3		5						
f	3				1		82	12	14	3		2	2				1			
θ	6	10	6		2	2	23	33	30	1	1	3	2						1	
s		3	2		1		3	5	93	2	2	5	3		1					
ʃ		1							3	107			8	1						
v				3		1	2	2	2		80	22	1						7	
z				1		2		2	3		27	73	1				1	3	3	4
tʃ		5	1			2			1	7			99	5						
dʒ					5	25						1	2	87						
m											9	5			82	7	10	2	2	3
n											6				17	75	20	1	1	
l								1			6	4	1		2	3	95	2	3	3
r							1	1	2		27	6			1		13	46	20	3
w								1			9	5					7	24	70	4
j			1					1		2	5	3		6	2	1	1	2	1	95

Table 4

Consonant confusion Matrix for hearing-impaired subjects, 8 sine waves quiet

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	M	n	l	r	w	j
p	76	12	32																	
t		83	10			3	1	1	3		1	3	14	1						
k	5	8	93			3	1	1					4	3	1	1				
b	6		1	68	7	13	3	3			16	3								
d			1	3	78	27		1			4	2		4						
g	1		1	1	9	93	1		1		4	2	1	6						
f	2	2					83	8	15			3	3	1	1		1		1	
θ	5	10	5	1		3	17	35	35	1		3	4						1	
s		3		2	2	1	4	3	96	1		4	4							
ʃ							1		4	106			9							
v				5	1	3	2	2			71	24		1				5	5	1
z	1			1		1		3	1	1	23	77			1			1	6	4
tʃ		3				4			3	4			102	4						
dʒ					3	24			1		1	3	2	86						
m						1		1			11	3			73	8	11	7	3	2
n											7		1		20	66	20	1	3	2
l							1	2			9	2	1		2	5	88	2	4	4
r							1	2	1		26	6			1	1	13	46	22	1
w								1			8	6			1		10	24	66	4
j	1	1						1		1	4	2		3	1	1	5	2	1	97

Table 5

Consonant confusion Matrix for hearing-impaired subjects. 4 sine waves quiet

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	64	11	3				1						1							
t	4	51	9			4	2	1	5				3	1						
k	3	4	50			1		4	4	1		2	5	5		1				
b	1			43	4	2	1	5			12	8				1			3	
d		1			30	22		1	2	1	5	10		3					5	
g			2	1		48			1			8	1	15			1	1	1	1
f	3					1	43	5	23	1		2	2							
θ	5	4	4	1		3	9	16	24		1	5	8							
s		1			1	1	6	5	51	3		6	4	1			1			
ʃ								3	1	68			8							
v				2	1	2	1	2	1		33	24	1				1		10	2
z				1		2		3	1		20	38					1	1	12	1
tʃ						4		1	2	4			65	4						
dʒ		1				22		1				2		53			1			
m											8				46	7	11	1	7	
n											5	3		1	15	22	18		13	3
l								1			4	1			4	4	58	1	6	1
r								1	1		14	3					15	14	31	1
w									1		9		1			1	11	10	45	2
j		1							1		8	1		1	1	1	9	2	14	41

Table 6

Consonant confusion Matrix for hearing-impaired subjects, unprocessed stimuli +10 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	42	13	48	1			10	1	2	2								1		
t	5	78	14			2	4	1	3	2			10							1
k	2	7	95	1					4	1		2	6	2						
b	3	1		51	21	11	3	2	2	1	17	2	1	3		1			1	
d				1	84	20			1		4	4	3	2					1	
g	1		1		9	90		1	2		3	3		6		1		1	1	1
f	22	6	1	1	1		59	8	15	2	3	1							1	
θ	8	12	3		4		34	27	21	3	2	4	1						1	
s		6				1	8	1	95	4		2	3							
ʃ		3				1	1		3	105			6							1
v				12	8	5	3	2			57	26		2				1	4	
z					5		1	1	5		15	88		2				1	2	
tʃ		5	1			1			5	3	1		101	3						
dʒ					3	26					2		1	88						
m								1			8	4			76	11	12	4	2	2
n	1			1	1			1			1	1			23	63	24	1	1	2
l				1				1			3	1			6	4	96	4	2	2
r								1	2	1	15	7					15	51	24	4
w						1	1	1			12	5			2	4	11	28	50	5
j		1			1	1					6	1	2	2	3	3	3	2	1	94

Table 7

Consonant confusion Matrix for hearing-impaired subjects, 16 sine waves +10 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	51	11	26	1			18	3	4	1	1	2	2							
t	3	77	12			1	6	1	8	1	1	2	6	2						
k	15	11	71			3	9	2	3	1	1		3	1						
b	5	1		55	10	5	7	1	6		22	4		1					3	
d	1	1	2	7	63	19					10	9	1	4					3	
g					6	81	1	1	5		5	7		11				1	2	
f	13	2	4	5	1		50	15	19		5	1	2	1				1	1	
θ	16	9	3	2	2	2	32	25	21	2	3	2	1							
s	5	8	1		1		4		87	3		5	4		2					
ʃ									7	101		1	11							
v				12	2	4		1	1		71	21		1					6	1
z				3		2		3	2		30	75		2					2	1
tʃ		3				6			3	6		1	100	1						
dʒ			1	1	3	21		1			1	1	3	87						1
m								2	1		8	5	2		68	11	7	11	5	
n								2			7			1	28	49	23	2	5	3
l								1			11	1	1		5	6	82	7	1	5
r		1		1	1			1	2		17	7				1	14	43	27	5
w									1		19	3		1		1	8	33	47	7
j				1	1		1	1			11			1	4	6	5	4	2	78

Table 8

Consonant confusion Matrix for hearing-impaired subjects, 12 sine waves +10 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	47	18	20	2	1		16	3	4	1	2	2	1	2					1	
t	3	66	7	1		1	13	7	7	1	1	1	10	1	1					
k	15	10	72		1	3	7	1	4	2			5							
b				36	8	9	8	2	3	1	41	5		1					3	3
d	1			5	63	21		2		2	5	11		6					4	
g		1		3	5	76	1	2	1	1	8	10		10					2	
f	14	2	6	3		2	54	12	16		3	1	4						2	1
θ	13	14	4	1	1	2	34	25	17	1	1	5	1					1		
s	6	2			1		8	3	90	2		7	1							
ʃ		1					1		4	99		1	14							
v		2	1	14	3	2		2	1		63	22		1				2	6	1
z				2				1	4		29	73	1		1	1	1		4	3
tʃ		3	1		1	4	1	2	3	6			96	3						
dʒ			1		1	22	1	1	1		1		3	86			1	1		1
m								1			18	3			74	9	7	6		2
n								3	1		7				36	41	26	2	1	3
l						1		1			11	1			7	8	74	6	6	5
r					1		1	1	2		26	6			1		14	41	24	3
w								1			17	6	1			1	7	34	47	6
j					1			2	2		14	3	1	2	6	2	6	3	4	74

Table 9

Consonant confusion Matrix for hearing-impaired subjects, 8 sine waves +10 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	38	12	24	1	1	3	25	2	9			1	3						1	
t	4	66	7	1		1	13	7	8		2	1	9						1	
k	18	12	67			1	6	2	7	3			2	2						
b	6	1		43	9	10	9	1	5		26	4		2					2	2
d	1			5	55	17		1	1	1	14	12	1	7		1			4	
g	1		1	3	8	69		3			11	9		11		1			3	
f	17	5	9	6		1	40	17	13	2	2	3	3			1			1	
θ	13	15	7	3	1	3	29	19	19	3	5		2						1	
s	3	5	1	1		1	7	7	87	3		3	1							1
ʃ								2	7	98		1	11	1						
v	1			12	1	4	2	4	1		60	26		2				2	5	
z				2	1	3		4	1		30	71					2	2	3	1
tʃ		2	1			2	1	1	6	4			96	5	1			1		
dʒ					2	24				2		2	3	85				1	1	
m					1	1		1			10	1			64	12	11	13	5	1
n				1				1			9	2	1		29	48	23	2	1	3
l					1	1			2		7	1			9	11	70	12	2	4
r						2		1	1		21	6			2		18	35	30	4
w							1	2			14	5		1	2		8	41	43	3
j	1							6	1		14	4		4	5	6	8	7	5	59

Table 10

Consonant confusion Matrix for hearing-impaired subjects, 4 sine waves +10 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	34	19	16				27	8	9	1	2		2	1				1		
t	4	60	5			1	16	3	15			2	13	1						
k	21	17	60		1	3	3	2	8	2			2					1		
b	4	4	2	30	18	7	13	4	3	1	17	11	1	2					3	
d	3			6	40	18	2	1	2		13	15		11	1	1			6	1
g				4	13	50		1	1	1	9	14	3	18					5	1
f	12	3	6	6	2	1	41	18	21	3	2	2	1			1			1	
θ	13	11	4	5	1	2	28	16	30		2	2	2		1				2	1
s	4	3	2	1	2		5	3	88	3		6	2						1	
ʃ		1	2		1	1	3		3	83		2	20	3			1			
v				11	4	3	2	1	2	2	54	25	1	2				1	10	2
z				4	3	3	1	2	2	1	42	38		1	1	2	2	3	10	5
tʃ	1	4	3			4	2		3	8			89	6						
dʒ					3	24				1		1	6	85						
m				1	1			1		1	16	4			43	9	18	4	16	6
n	1			1							4	4			26	23	35	6	12	8
l	1				1	1		1			8	2			7	7	57	12	15	8
r					1	1				2	25	7		1			26	18	33	6
w								1	1		25	4	1		3		14	22	41	8
j							1	4			15	2	2	4	5	6	20	12	14	35

Table 11

Consonant confusion Matrix for hearing-impaired subjects, unprocessed stimuli, 0 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	24	18	29	1	3	4	17	7	4	1	2	3	6	1						
t	4	49	31		1	1	5	3	7	1		1	12	5						
k	8	10	72	1	4	1	3	7	4	3			4	3						
b	11	4	6	27	14	20	2	2	1	9	10	1		5	3	2	2		1	
d	2		2	5	48	37	1	1	1	2	1	3	2	12		1				2
g	1	4	4		20	68		1	2	2	2	7	1	7						1
f	14	21	14	8	1		22	6	11	11	1	2	5				1		1	2
θ	8	18	5	6	8	7	13	11	19	7	5	6	4	2	1					
s	6	7	8	2	4		6	2	58	8		7	8	2					1	1
ʃ	1	1	3			2		1	4	86		1	18	1			1			1
v	3	1	4	19	9	8	3	5	2	4	26	15	3	5		1	3	1	4	4
z	2			4	4	9		1	1	1	24	48	1	2		1	1	2	4	15
tʃ		8	4			4	1	1	1	11		1	85	3		1				
dʒ		1	1		5	22	1	1	2	2			8	74						3
m	1			1	5	4		1			10	6		1	43	13	8	2	6	19
n					2	4	1				2	1		3	26	33	30	3	2	13
l	1		3	8	3	2	1	2	2		10	3		5	17	11	33	6	3	10
r	1	1	2		7	3			2	3	21	11	2	5	2	3	11	15	13	18
w		2	2	1	5	1	3	2	3	2	16	8	1	2	5	4	10	13	21	19
j					1	2				1	8	4	3	8	8	6	12	2	6	59

Table 12

Consonant confusion Matrix for hearing-impaired subjects, 16 sine waves, 0 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	34	17	23	2	3		19	2	7	3	1	2	4	1				1		1
t	5	46	20	1	1	1	13	2	7	6		2	13	3						
k	10	14	40	1	6	8	7	5	9	3	1	5	5	3		1				2
b	4	1	2	25	22	14	6	4	3	1	25	5	1	2				1	1	3
d	2	2	1	10	44	26	1	4		1	10	8	3	4	1			1		2
g	1	4	3	3	15	58			1	1	4	10	2	14		1				3
f	13	14	22	8	7	4	25	5	6	2	3	3	2	1		1		1		3
θ	6	11	5	12	4	5	15	9	18	4	10	9	3	4		1		3		1
s	8	22	9		1	5	7	5	43	3	2	5	5	2	1					2
ʃ		1	9						11	65		1	28	5						
v	3	4		22	10	11	2	2	1		29	22	2	2	2				2	6
z	2			5	11	10	1	1	1	1	22	40	1	7	1	2		2	3	10
tʃ		9	11			8		2	2	19		1	62	5					1	
dʒ	3	2	5	1	8	25		2	2		1	3	12	53	1				1	1
m			3	1	1	2					12	7	1	2	46	13	16	6	3	7
n	1	1	1	1	1	2		2	1		4	6	1	2	19	26	17	5	8	22
l	2	1	1	3	6	1	1				15	6		5	18	14	21	6	2	18
r		1	1	3	5	7	2	3			35	15		3		2	1	11	16	15
w	1	1	3	1	3	4	2	2	2	2	20	10	1		12	1	10	9	19	17
j	1	1	4	1	2	5		1	2	1	14	4		6	4	7	7	5	7	48

Table 13

Consonant confusion Matrix for hearing-impaired subjects, 12 sine waves, 0 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	25	20	12	6	2	2	24	6	10	3		1	5	2	1		1			
t	11	35	15	1	8	1	6	4	10	1		1	14	10					1	2
k	16	22	30		5	11	8	3	10	5		2	4	2		1		1		
b	7	3	10	26	12	9	10	6	4	1	17	4	1	7		2				1
d	1	2	1	5	43	28		1	3		4	13	3	9					1	6
g	2		3	6	20	44	2	1	3		8	10	1	10		2	1	1	2	4
f	13	5	8	10	16	4	30	3	10	5	5	4	5	1						1
θ	14	19	8	8	7	8	9	7	22	1	1	5	4	5	1					1
s	3	9	12	1	3	5	9	7	52	3	2	7	6							1
ʃ	1	5	6	2	3	4		2	9	44	1	2	38	3						
v	2	3	2	10	18	13	2	2	2		30	23	1	3	1			3	3	2
z	4		2	3	10	8		1	1		30	35	1	4	2	1		4	7	7
tʃ	2	6	13			5	1	1	5	5		1	77	4						
dʒ	1	2	7		5	31	1	1	1	1	1	2	5	59					1	2
m	2		1	1	2			1		1	11	1		3	45	15	8	6	1	22
n	1			2			1	2		1	12	1	1	2	23	23	20	12	6	13
l	1	1	1	6	3	1		1	1	1	17	7		4	14	14	20	10	5	13
r	3			4	5	3	2		1	2	33	15		3		1	7	13	15	13
w	2	1		2	3	6	2		1		15	12	1	6	6	5	8	15	19	16
j	2	2	2	1	4	1	1	1	2		19	8		3	11	5	8	11	8	31

Table 14

Consonant confusion Matrix for hearing-impaired subjects, 8 sine waves, 0 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	22	19	20	1	2	3	32	5	5	1		3	4				1	1		1
t	10	44	12		7	1	9	5	13		2	2	10	5						
k	17	19	38	2	2	1	8	1	8	3	1	1	12	5		1				1
b	8	1	2	22	19	10	14	7	3	2	17	5	3	3	1		1			2
d	3	1	1	6	42	19	1	1	6		9	13		8	2	3	1	2		2
g	3	3	1	7	20	41	1	2	2	2	7	8	1	15		1			1	5
f	10	10	5	17	6	7	29	6	11	1	3	1	6	3	1				1	3
θ	13	12	5	6	12	4	26	6	15	4	5	1	6	1		1		1		2
s	4	10	9	8	6		9	5	52	4	1	5	3	1	1	1				1
ʃ		7	11		4	4		1	13	42		1	29	7		1				
v	3	3	1	6	16	11	6	2	8	1	34	15	1	3	1	1	2	2	3	1
z	4		2	4	20	9	2	1	1		22	25		13	1	1	1		6	8
tʃ	1	9	8	1	1	7		1	3	7			73	9						
dʒ	1	1	5	2	12	19	1	3	2	3		5	5	55		1			1	4
m	1			1							10	4		1	44	13	13	11	8	14
n			1	1					1		9	6			28	21	23	5	9	16
l	2		1	2	6	3	1		1	2	14	6		3	14	6	35	6	10	8
r	1	2	2	3	5	5	1	2	3	1	20	16		2	1	4	8	17	16	11
w	2	1		5	7	6	1		1		16	6	3	4	9	4	8	6	21	20
j	1	1	1	2	3	3			2	2	17	9		5	9	7	14	3	12	29

Table 15

Consonant confusion Matrix for hearing-impaired subjects, 4 sine waves, 0 dB SNR

	p	t	k	b	d	g	f	θ	s	ʃ	v	z	tʃ	dʒ	m	n	l	r	w	j
p	21	24	15	3	1	3	9	5	11	8		7	6	2	1		1		3	
t	13	22	16	6	9	4	15	3	14	3	1	2	6	3			1			2
k	18	13	28	3	9	5	15	1	10	4		1	8	3	1			1		
b	5	2	3	22	29	15	8	3	8	1	6	2	2	7	1	1		1	3	1
d	4	1		14	34	26		4	7		6	9		6				1	6	2
g	5	3	6	2	24	29	5	2	2	2	6	13	4	11			1		4	1
f	15	21	9	4	5	2	18	6	21	6		4		2			1	3	2	1
θ	12	16	11	3	5	4	16	7	22	3	5	3	3	2		2			4	2
s	9	27	12	3	2	1	2	5	32	4		10	6	4	1	1		1		
ʃ	4	13	12	2	7	12	4	3	7	9		6	26	13	1		1			
v	2	2	2	10	21	15	1	2	2	3	17	14	3	10	1	1	2	3	7	2
z	1			1	18	7		4	2	1	19	31	2	5		5	1	2	12	9
tʃ	4	27	6	1	6	8	5	4	7	3		4	33	10			1			1
dʒ		6		1	10	27		3	6	3		1	5	50					4	4
m	1	1		1	3			1	1		13	4	2	3	30	15	9	2	11	23
n	1	2	1		3					1	7	3		2	18	19	11	3	13	36
l		1	1	3	4	1	1	1	2	2	14	5		4	15	9	15	8	15	19
r	2	1	1	2	5	6			3		11	12		5	5	2	20	10	18	17
w	2	1	2	5	5	3	2	3	3	1	17	10	2	2	4	5	5	8	20	20
j	1	1	1		4	3		5	1	2	14	9		5	8	8	4	16	30	

REFERENCES

- Alcantara, J.I., Dooley, G.J., Blamey, P.J., Seligman, P.M. (1994). Preliminary evaluation of a formant enhancement algorithm on the perception of speech in noise for normally hearing listeners. Audiology. 33(1), 15-27.
- Allen, J.B. (1977). Short term spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform. IEEE Transactions on Acoustics, Speech and Signal Processing. ASSR(24) No.3, 235-238.
- ANSI (R2003) ANSI S3.1-1999 American National Standards, Maximum permissible ambient noise levels for audiometric test rooms. New York: American National Standards Institute Inc.
- ANSI (R1997) ANSI S3.5-1997. American National Standards, Methods for calculation of the Speech Intelligibility Index. New York: American National Standards Institute Inc.
- ASHA (1977). Manual Pure Tone Audiometry. ASHA committee on audiometric evaluation. Audiology Desk Reference Vol(2), 221-225
- Baer, T., Moore, B.C.J. & Gatehouse, S. (1993). Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing-impaired: Effects on intelligibility, quality and response times. Journal of Rehabilitation Research and Development. 30(1), 49-72.
- Baer, T., Moore, B. C. J, & Kluk, K. (2002) Effects of low pass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies. Journal of the Acoustical Society of America 112(3 Pt 1), 1133-44
- Balachandran, R., Levitt, H. (1997). Sine wave modeling of speech for normal hearing and hearing-impaired listeners. Presented at the National Institutes of Health, II conference on Hearing aid research and Development.

- Bench, J., & Bamford, J. (1979). Speech-hearing tests and the language of hearing-impaired children. London: Academic Press.
- Bilger, R.C., & Wang, M.D. (1976). Consonant confusions in patients with sensorineural hearing loss. Journal of Speech, Hearing and Language Research, 19, 718-748.
- Bochner, J. H., Garrison, W. M. , Sussman, J. E. , & Burkard, R. F. (2003). Development of materials for the clinical assessment of speech recognition: the speech sound pattern discrimination test. Journal of Speech, Language and Hearing research . 46(4), 889-900.
- Boothroyd, A. (1968). Developments in speech audiometry. Sound, 2, 3-10.
- Boothroyd, A. Mulhearn, B., Gong, J., Ostroff, J. (1996). Effects of spectral smearing on phoneme and word recognition. Journal of the Acoustical Society of America, 100(3), 1807-1818.
- Borden, G.J., Harris, K. S., & Raphael, L.J. (1994). Speech Science Primer: Physiology, Acoustics and Perception of Speech. (3rd Edition). Baltimore: Williams and Wilkins.
- Bunnell, T. (1990). On enhancement of spectral contrast in speech for hearing-impaired listeners. Journal of the Acoustical Society of America, 88, 2546-2556.
- Byrne, D., & Dillon, H. (1986). The National Acoustic Laboratories (NAL) New Procedure for selecting the Gain and Frequency Response of A Hearing Aid. Ear & Hearing, 7(4), 257-265.
- Cazals, Y., Palis, L. (1991). Effect of silence duration in intervocalic velar plosive on voicing perception for normal and hearing-impaired subjects. Journal of the Acoustical Society of America , 89(6), 2916-2922.

- Ching, Y. C., Dillon, H., & Byrne, D. (1998). Speech recognition of hearing-impaired listeners: Predictions from audibility and limited role of high frequency amplification. Journal of the Acoustical Society of America, 103(2), 1128-1140.
- Ching, Y. C., Dillon, H., Katch, R., & Byrne, D. (2001). Maximizing effective audibility in hearing aid fitting. Ear and Hearing, 22(3), 212-224.
- Dillon, H. (1997). Converting Insertion Gain to and from headphone Coupler Responses. Journal of speech and Hearing Research, 23, 643-669.
- Dorman, M.F., Lindholm, J.M., & Hanneley, M.T. (1985). Influence of the first formant on the recognition of voiced stop consonants by hearing-impaired listeners. Journal of Speech, Hearing and Language Research, 28, 377-380.
- Dorman, M., Loizou, P. (1998). The identification of consonants and vowels by cochlear implant patients using a 6 channel continuous interleave sampling processor and by normal hearing subjects using simulations of processors with 2 to 9 channels. Ear and Hearing, 19(2), 162-166.
- Dorman, M., Loizou, P., & Rainey, D. (1997). Speech Intelligibility as a function of the number of channels of stimulation for signal processors using sine wave and noise-band outputs. Journal of the Acoustical Society of America, 102, 2403-2411.
- Dorman, M., Loizou, P., Fitzke, J., & Tu, Z. (1998). The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant speech processors with 6-20 channels. Journal of the Acoustical Society of America, 104, 3583-3585.
- Dorman, M., Loizou, P., Spahr, A.J., & Maloff, E. (2002). A comparison of speech understanding provided by Acoustic Models of Fixed-channel and channel-picking signal processors for cochlear implants. Journal of Speech, Language and Hearing Research, 45, 783-788.

- Dubno, J. R., & Levitt, H. (1981). Predicting consonant confusions from acoustical analysis. Journal of the Acoustical Society of America, 69(1), 249-261.
- Dubno, J.R., Dirks, D.D., & Langhofer, L.R. (1982). Evaluation of hearing-impaired listeners using a nonsense syllable test. II : Syllable recognition and Consonant confusion patterns. Journal of Speech, Hearing and Language Research, 25, 141-148.
- Dubno, J.R., Dirks, D.D., & Schaefer, A. B. (1987). Effects of hearing loss on utilization of short-term spectral cues in plosive recognition. Journal of the Acoustical Society of America, 81(6), 1666-1675.
- Dubno, J. R., Dirks, D. D., & Ellison, D. E. (1989). Stop-consonant recognition for normal-hearing and listeners with high-frequency sensorineural hearing loss: The contributions of selected frequency regions. Journal of the Acoustical Society of America, 85(1), 347-355.
- Florentine, M., Buus, S., Scharf, B., & Zwicker, E. (1980). Frequency selectivity in normal-hearing and hearing-impaired observers. Ear and Hearing, Vol.18(4), 346-348
- Florentine, M., & Buus, S. (1984). Temporal Gap detection in sensorineural hearing impaired listeners. Journal of Speech, Language and Hearing research, 27(3), 449-455.
- Fu, Q.-J., Shannon, R., & Wang, X. (1998). Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and Electric hearing. Journal of the Acoustical Society of America, 104, 3586-3596.
- Glasberg, B. R., & Moore, B. C. J. (1986). Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. Journal of the Acoustical Society of America, 79(4), 1020-1033.

- Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched noise data. Hearing Research, 47(1-2), 103-38.
- Hillenbrand, J., Getty, L., Clark, M., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. Journal of the Acoustical Society of America, 97, 3099-3111.
- Johnson, D., Whaley, P., & Dorman, M. (1984). Processing cues for plosive voicing by young hearing-impaired listeners. Journal of Speech, Hearing and Language Research, 27, 112-118.
- Kates, J. M. (1991). A simplified representation of speech for the hearing-impaired. Journal of the Acoustical Society of America, 89, 4(Pt 2), 1961.
- Kates, J.M. (1994). Speech Enhancement Based on a Sine Wave Model. Journal of Speech and Hearing Research, 37, 449-464.
- Kennedy, E., Levitt, H., Neuman, A.C., & Weiss, M. (1998). Consonant/Vowel intensity ratios for maximizing consonant recognition by hearing-impaired listeners. Journal of the Acoustical Society of America, 103(2), 1098-1114.
- Kochkin, S (2002). MarkeTrak VI: Hearing Aid Industry Market Tracking Survey 1984-2000. Knowles Electronics, Inc.
- Klatt, D.H. (1980). Software for cascade/parallel formant synthesizer. Journal of the Acoustical Society of America, 48(2) 971-995.
- Kryter, K.D. (1956). On predicting the intelligibility of speech from acoustical measures. Journal of Speech and Hearing Disorders, 21 (2), 208-217.

- Leek, M.R., Dorman, M.F., & Summerfield, Q. (1987). Minimum spectral contrast for vowel identification by normal hearing and hearing-impaired listeners. Journal of the Acoustical Society of America, 81 (1), 148-154.
- Leek, M.R., & Summers, V. (1996). Reduced frequency selectivity and the preservation of spectral contrast in noise. Journal of the Acoustical Society of America, 100, 1796-1806.
- Leek, M.R., & Van Summers (1993). Auditory filter shapes of normal-hearing and hearing-impaired listeners in continuous broadband noise. Journal of the Acoustical Society of America, 94, 440-449.
- Levitt, H. (1993). Future directions in hearing aid research. Journal of Speech-Language Pathology and Audiology, Revue d'orthophonie et d'audiologie. Monograph Supplement, 1.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. Journal of the Acoustical Society of America 49(2), Suppl 2, 467-473
- Levitt, H., Bakke, M., Kates, J., Neuman, A., Schwander, T., & Weiss, M. (1993). Noise reduction strategies for hearing aids. Scandinavian Audiology, Suppl.(38), 7-19.
- Liberman, A. M., Delattre, P.C., Cooper, F.S., & Gerstmann, L.J. (1954). The role of consonant vowel transitions in the perception of stop and nasal consonants. Psychology Monographs. Gen Appl. 68, 1-13.
- Lindholm, J.M., Dorman, M., Taylor, B.E., & Hannley, M. T. (1988). Stimulus factors influencing the identification of plosiveplosives by normal-hearing and hearing-impaired adults. Journal of the Acoustical Society of America, 83(4), 1608-1615.
- Loizou, P.C., Dorman, M.F., & Zhemin, T. (1999). On the number of channels needed to understand speech. Journal of the Acoustical Society of America, 106, 2097-2103.

- Loizou, P.C., Dorman, M.F., & Tu, Z. (1999). On the number of channels needed to understand Speech. Journal of the Acoustical Society of America, 106(4 Pt 1), 2097 -103.
- Loizou, P.C., Dorman, M.F., Tu, Z., & Fitzke, J.(2000). The recognitions of sentences in noise by normal-hearing listeners using simulations of SPEAK-type cochlear implant processors. Annals of Otology, Rhinology and Laryngology, 109(12), Suppl.185, 67-68.
- Loizou, P.C., Dorman, M.F., Spahr, A.J., & Maloff, E.(2002). A comparison of speech understanding provided by acoustic models of fixed channel and channel picking processors for cochlear implants. Journal of the Acoustical Society of America, 45(4) 783-788.
- MacLeod, A., & Summerfield, Q. A. (1990). A procedure for measuring auditory and audio-visual speech reception thresholds for sentences in noise: rationale, evaluations, and recommendations for use. British Journal of Audiology, 24, 29-43.
- Mc Aulay, R.J. & Quatieri, T.F.(1986). Speech Analysis/synthesis based on a sinusoidal representation . IEEE Transactions on acoustics, Speech and Signal processing, 34, 744-754.
- Markel, J. D., & Gray, A. H. (1976). Linear Prediction of Speech. New York: Springer.
- Miller, G. & Nicely, P. (1955) An analysis of perceptual confusions among English consonants. Journal of the Acoustical Society of America, 27(2), 338-352.
- Montgomery, A. A., & Edge, R. A. (1988) Evaluation of two speech enhancement techniques to improve speech intelligibility. Journal of Speech, Hearing and Language Research, 31, 386-393.

- Moore, B. C. J., & Glasberg, B. R. (1983). Suggested formulae for calculating auditory filter bandwidths and excitation patterns. Journal of the Acoustical Society of America, 74, 750-753.
- Moore, B. C. J., & Glasberg, B. R. (1986). Comparison of frequency selectivity in simultaneous and forward masking for subjects with unilateral cochlear impairment. Journal of the Acoustical Society of America, 80, 93-107.
- Moore, B. C. J., & Glasberg, B. R. (1998). Use of a loudness model for hearing aid fitting. I: Linear hearing aids. British journal of Audiology, 32(5), 317-335.
- Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R., & Alcantara, J. I. (2000). A test for the diagnosis of dead regions in the cochlea. British Journal of Audiology, 34(4), 205-224.
- Nakisa, M. (2000). Consequences of varying the perceptual parameters of cochlear-implants. Unpublished Doctoral dissertation, University College London, London
- Needleman, A.R., & Crandell, C.C. (1995). Speech recognition in noise by hearing-impaired and noise-masked normal-hearing listeners. Journal of the American Academy of Audiology 6(6), 414-24.
- Nilsson, M., Soli, S., & Sullivan, J. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and noise. Journal of the Acoustical Society of America, 95, 1085-1099.
- Olsen, W., & Noffsinger, D. (1974). Comparison of one new and three old tests of auditory adaptation. Archives of Otolaryngology, 99, 94-99.

- Owens, E., Benedict, M., & Schubert, E. (1972). Consonant phonemic errors associated with pure-tone configurations and certain kinds of hearing impairment. Journal of Speech, Hearing and Language Research, 15, 308-322.
- Plomp, R. (1994). Noise, amplification and compression: considerations of three main issues in hearing aid design. Ear and Hearing, 15, 2-12.
- Quatieri, T.F., & McAulay, R. J. (1986). Speech transformations based on a Sine Wave model. IEEE Transactions on Acoustics, Speech and Signal processing, 34, 1449-1464.
- Raz, I., & Noffsinger, D. (1985). Identification of synthetic, voiced stop-consonants by hearing - impaired listeners. Audiology, 24, 437-448.
- Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. Science, Vol. 212, 947-950.
- Remez, R.E., & Rubin, P.E. (1990). On the perception of speech from time-varying acoustic information: Contributions of amplitude variation. Perception and Psychophysics, 48(4), 437-445.
- Resnick, S.B., Dubno, J.R., Hoffnung, S., & Levitt, H. (1975). Phoneme errors on a nonsense syllable test. Journal of the Acoustical Society of America, Suppl(1), 58, S114.
- Revoille, S., Pickett, J. M., Holden-Pitt, L. D. (1985). Perceptual cues to the voiced-voiceless distinction of final fricatives for listeners with impaired or with normal-hearing. Journal of the Acoustical Society of America, 77(3), 1263-1265.
- Revoille, S., Pickett, J. M., Holden-Pitt, L.D., Talkin, D. (1982). Acoustic cues to final stop voicing for impaired and normal-hearing listeners. Journal of the Acoustical Society of America, 72(4), 1145-1154.

- Revoille, S., Pickett, J. M., Holden-Pitt, L.D., Talkin, D., & Brandt, F. D. (1987). Burst and transition cues to voicing perception for spoken initial plosives by impaired and normal hearing listeners. Journal of Speech, Hearing and Language Research, 30, 3-12.
- Rosen, S. (1992). Temporal information in speech acoustic, auditory and linguistic aspects. Philosophical Transaction of the Royal Society of London Series B, Biological Sciences. 367-373.
- Rosen, S., Faulkner, A., & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. Journal of the Acoustical Society of America, 106(6), 3629-3636.
- Schum, D. J., Matthews, L.J., & Lee, F. S. (1991). Actual and predicted word-recognition performance of elderly hearing-impaired listeners. Journal of Speech, Language and Hearing research, 34(3), 636-642.
- Shannon, R., Zeng, F.G., & Wygonski, J. (1998). Speech recognition with altered spectral distribution of envelope cues. Journal of the Acoustical Society of America, 104, 2467-2476.
- Shannon, R., Zeng, F.G., & Wygonski, J., Kamath, V., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. Science, 270, 303-304.
- Simpson, A. M., Moore, B. C. J., & Glasberg, B. R. (1990). Spectral Enhancement to improve the intelligibility of speech in noise for hearing-impaired listeners. Scandinavian Otolaryngology, Suppl(469), 101-107.
- Skinner, M. W., Clark, G. M., Whitford, L. A., Seligman, P. M., Staller, S. J., Shipp, D. B., Shalloo, J. K., Everingham, C., Menapace, C. M., Arndt, P. L., Antogenelli, T., Brimacombe, J. A, Pijl, S., Daniels, P., George, C. R, McDermott, J.H., & Beiter, A. L.

- (1994). Evaluation of a new spectral peak coding strategy for the Nucleus 22 channel cochlear implant system. The American Journal of Otology, Vol. 15(2), 15-27.
- Stone, M. A., & Moore, B. C. J. (1992). Spectral feature enhancement for people with sensorineural hearing-impaired: Effects of speech intelligibility and quality. Journal of Rehabilitation Research and Development, 29, 35-56.
- Summerfield, Q., Foster, J., & Tyler, R. (1985). Influences of formant bandwidth and auditory frequency selectivity on identification of place of articulation in plosives. Speech Communication, 4, 213-229.
- Turner, C. W., & Cummings, K. J. (1999). Speech audibility for listeners with high-frequency hearing loss. American Journal of Audiology, 8, 1059-0889.
- Turner, C.W., & Robb, M.P. (1987). Audibility and recognition of plosives in normal-hearing and hearing-impaired subjects. Journal of the Acoustical Society of America, 81(5), 1566-1573.
- Turner, C. W., & Henry, B. A. (2003). Benefits of amplification for speech recognition in background noise. Journal of the Acoustical Society of America, 112(4), 1675-1680.
- Vickers, D. A., Moore, B. C. J., & Baer, T. (2001). Effects of low-pass filtering on the intelligibility of speech in quiet for people with and without dead regions at high frequencies. Journal of the Acoustical Society of America, 110(2), 1164-1175.
- Van Tassel, D. J., Hagen, L. T., Koblas, L. L., & Penner, S. G. (1982). Perception of short-term spectral cues for plosive place by normal hearing and hearing-impaired subjects. Journal of the Acoustical Society of America, 72(6), 1771-1780.

- Walden, B. E., & Montgomery, A. A. (1975). Dimensions of consonant perception in normal and hearing-impaired listeners. Journal of Speech, Hearing and Language Research, 18, 444-455.
- Wang, M., Reed, C. M., & Bilger, R. C. (1978). A comparison of filtering and sensorineural hearing loss on patterns of consonant confusions. Journal of Speech, Hearing and Language Research, 21, 5-36.
- Wilson, B.S., Lawson, D.T., Zerbi, M. (1995). Advances in coding strategies for cochlear implants. Advances in Otolaryngology -Head and Neck Surgery , Vol.9, 105-129.
- Winer, B. J. (1971). Statistical principles in experimental design (2nd Ed) New York: McGraw-Hill, 198-199.
- Zeng, F. G., & Turner, C. W. (1990). Recognition of voiceless fricatives by normal and hearing-impaired subjects. Journal of Speech, Hearing and Language Research, 33, 440-449.