Analysis of Frequency-Smearing Models Simulating Hearing Loss

by

Alan T. Asbeck

Submitted to the Department of Electrical Engineering and Computer Science on May 21, 2003, in partial fulfillment of the requirements for the degree of Master of Engineering in Electrical Engineering and Computer Science

Abstract

Various authors have created models of the psychoacoustic effects of sensorineural hearing loss that transform ordinary sounds into sounds that evoke the perception of hearing loss in normal-hearing listeners. In this thesis, models of the reduced frequency resolution, reduced temporal resolution, and loudness recruitment and absolute threshold loss were evaluating using a model of human perception. Confusion matrices of vowels were simulated using these models and compared to confusion matrices from five hearing-impaired subjects. It was found that the model of human perception used spectral cues in a different way than actually occurs in humans, making it difficult to evaluate the hearing loss models, and causing the simulated confusion matrices to be different than the real subjects' matrices. Even so, the frequency smearing models caused the total error rate to increase with increasing smearing bandwidth, and the results were generally consistent with the expected behavior.

Thesis Supervisor: Louis D. Braida
Title: Henry Ellis Warren Professor of Electrical Engineering
Acknowledgments

I first want to thank my advisor, Professor Louis Braida, for his help and support throughout this project, and for giving me a project in the first place. I especially want to thank him for putting up with my tendency to procrastinate, and for spending so much time reading my thesis and being there at the end.

I want to thank Michael “Q” Chin, for really making my time here a wonderful experience, and for helping me think through this project and other things throughout the year. I’ve really appreciated the time we’ve been able to get to know each other, including everything from the “sea-snakes” on the chalkboard and talking about the various topics you have come up with.

I also want to thank Peninah Rosengard, Charlotte Reed, Andrew Oxenham, and everyone else in the Sensory Communication group that has helped me along the way.

Finally, I acknowledge the NIH and NSF, which have provided funding for my work on this thesis.
Contents

1 Introduction 25
  1.1 Background ........................................ 25
  1.2 Physiology of Hearing Impairment .................. 25
  1.3 Psychoacoustics of Hearing Impairment .......... 28
  1.4 Simulating Hearing Loss ........................... 31
    1.4.1 Reasons for Simulating Hearing Loss .......... 31
    1.4.2 Previous Work Simulating Hearing Loss ....... 32
  1.5 Problem Statement .................................. 35

2 Signal Processing in the Experiment 36
  2.1 Overview of Experiment ............................. 36
  2.2 Overview of Signal Processing ...................... 37
    2.2.1 Processing by Real Subjects ................... 37
    2.2.2 Processing by the Simulation .................. 37
  2.3 Detail on the Brantley Confusion Matrix Experiment ........... 40
  2.4 Speech Stimuli and Vowel Extraction ............... 41
  2.5 Signal processing simulating hearing loss .......... 42
    2.5.1 Methods of frequency smearing ................. 42
    2.5.2 Method of Recruitment .......................... 61
    2.5.3 Combining the Smearing Methods and Recruitment .... 71
  2.6 $d'$ Discrimination Model .......................... 71
  2.7 Decision Model ..................................... 75
3 Results

3.1 Total Error Rates ........................................... 78

3.2 Locations of Errors in Confusion Matrices ..................... 79
  3.2.1 Comparison of Moore and ter Keurs Methods ................. 84

3.3 Effects of Processing Parameters on the Results ............... 85
  3.3.1 Loudness recruitment and Audibility ....................... 85
  3.3.2 Male vs. Female Speakers ................................. 86

3.4 Results with Simulating Full Audibility ....................... 87
  3.4.1 Error rates ............................................. 88
  3.4.2 Locations of Errors ..................................... 89
  3.4.3 Comparison of Moore and ter Keurs Methods ................ 89

4 Discussion .................................................. 118

4.1 Why the Simulation Did Not Work .............................. 118

4.2 Other Sources of Error ....................................... 122

4.3 Conclusions that Can Be Drawn ................................ 122
  4.3.1 Total Error Rates and Smearing and Recruitment .......... 123
  4.3.2 Comparing Moore and ter Keurs Smearing .................. 126
  4.3.3 The Effect of Audibility .................................. 127

4.4 Potential Perceptual Models in Humans ......................... 128

5 Conclusions .................................................. 132

5.1 Summary of Findings ........................................ 132

5.2 Relevance to Hearing Aid Signal Processing .................... 133

5.3 Recommendations for Future Work .............................. 133

A Audiograms and Confusion Matrices for the Real Subjects .... 135

B Tables of Computed Confusion Matrices .......................... 140

C Plots of $\chi^2$ Values ....................................... 151
D Recommendations for an Improved Simulation 183
D.1 Rationale and Background ....................................... 183
D.2 Overview of Model .............................................. 183
D.3 Details of Model ................................................ 184
D.4 Simulating Hearing Loss ........................................ 186
D.5 Possible Sources of Error ....................................... 187

References 189
List of Figures

1-1 Figure showing how the basilar membrane is compressive for tones at the best frequency of the location of measurement. This figure shows “the level of a masker required to mask [a] 6-kHz [tone] signal, as a function of signal level” (Oxenham and Plack, 1997). The plot of the on-frequency masker is linear, as would be expected. However, the off-frequency masker shows compression, indicating both that the basilar membrane is compressive and that it responds linearly for tones an octave below the best frequency of a given location. Taken from Oxenham and Plack, 1997.

1-2 Loudness matching curves for subjects with unilateral hearing loss. The thin line at a 45-degree angle is the expected result if the two ears gave equal loudness results for a tone at a given level. The fact that the actual data is at a steeper slope than this indicates that for the impaired ear, as the stimulus sound pressure level increases a little, the apparent loudness increases more rapidly than in a normal ear. Also, the asterisks indicate the subjects’ absolute thresholds; these are elevated in the impaired ear relative to the normal ear. Taken from Moore et al., 1996, with the 45-degree line added.

2-1 Signal processing overview.
2-2 Block diagrams of Moore frequency smearing algorithm. The top block diagram shows the main processing sequence, and the bottom block diagram shows detail for the “Smear Spectrum” block in the top part. Taken from Baer and Moore, 1993.

2-3 Spectra resulting from different amounts of Moore smearing on the vowel II taken from the word BIIDH. All files were put through the recruitment algorithm but with no recruitment. The unsmeared signal is shown at the regular level, while the other traces have been offset by -40 dB increments. The labels to the left of the traces indicate the amount of smearing, in terms of the effective number of ERBs of the widened auditory filters; US indicates the unsmeared signal.

2-4 Channels resulting from different amounts of Moore smearing on the vowel II taken from the word BIIDH. All files were put through the recruitment algorithm but with no recruitment. The unsmeared signal is shown at the regular level, while the other traces have been offset by -8 dB increments. The labels to the left of the traces indicate the amount of smearing, in terms of the effective number of ERBs of the widened auditory filters; US indicates the unsmeared signal.

2-5 Sample spectra showing the results of smearing, for a synthetic vowel /ae/. The different frames (vertical dimension in the figure) show successive chunks of signal that are partially overlapping. As can be seen in the right column, showing the output spectra, the effective degree of smearing is less than was intended. The intended degree of smearing is seen in the middle column, which shows the results after smearing but before the overlap-add process. Taken from Baer and Moore, 1993.

2-6 Block diagram showing the details of the ter Keurs processing method. Taken from ter Keurs et al., 1992.
2-7 Block diagram showing the details of the Spectral Smearing block from 2-6. Taken from ter Keurs et al., 1992. .......................... 53

2-8 Sample of the outputs of the ter Keurs frequency smearing algorithm. The top log spectrum, labeled “US,” shows the unsmeared signal, while the bottom three spectra show the results of smearing of 1/8, 1/2, and 2 octaves, as labeled. Taken from ter Keurs et al., 1992. .............. 54

2-9 Spectra resulting from different amounts of ter Keurs smearing on the vowel II taken from the word BIIDH. All files were put through the recruitment algorithm but with no recruitment. The unsmeared signal is shown at the regular level, while the other traces have been offset by -50 dB increments. The labels to the left of the traces indicate the amount of smearing, in octaves; US indicates the unsmeared signal. One octave of smearing is approximately equivalent to 3.9 ERBs of smearing. ......................................................... 55

2-10 Channels resulting from different amounts of ter Keurs smearing on the vowel II taken from the word BIIDH. All files were put through the recruitment algorithm but with no recruitment. The unsmeared signal is shown at the regular level, while the other traces have been offset by -12 dB increments. The labels to the left of the traces indicate the amount of smearing, in octaves; US indicates the unsmeared signal. One octave of smearing is approximately equivalent to 3.9 ERBs of smearing. ......................................................... 56

2-11 Block diagram of perceptual model of human temporal processing, used by Hou and Pavlovic in providing rationale for their temporal smearing model. Taken from Hou and Pavlovic, 1994. ......................... 58

2-12 Block diagram of the Hou temporal smearing model. The top diagram (a) shows the overall signal processing path, and the bottom diagram (b) shows the specifics of how temporal smearing is accomplished. Figure reproduced from Hou and Pavlovic, 1994. ......................... 59
2-13 Sample of outputs from the Hou temporal smearing model. The top panels (a and b) show the unsmeared signal and its spectrum, and the lower panels show the signal and its spectrum after it has been smeared with an ERD of 0.1ms and 28ms, respectively. Figure taken from Hou and Pavlovic, 1994.

2-14 Spectra resulting from different amounts of Hou smearing on the vowel II taken from the word BIIDH. All files were put through the recruitment algorithm but with no recruitment. The unsmeared signal is shown at the regular level, while the other traces have been offset by -50 dB increments. The labels to the left of the traces indicate the amount of smearing, that is, the equivalent rectangular duration of the smearing temporal window, in msec; US indicates the unsmeared signal.

2-15 Channels resulting from different amounts of Hou smearing on the vowel II taken from the word BIIDH. All files were put through the recruitment algorithm but with no recruitment. The unsmeared signal is shown at the regular level, while the other traces have been offset by -8 dB increments. The labels to the left of the traces indicate the amount of smearing, that is, the equivalent rectangular duration of the smearing temporal window, in msec; US indicates the unsmeared signal.

2-16 Block diagram of the model of loudness recruitment by Moore and Glasberg, 1993. The parallel arrows between the middle blocks represent the outputs of the 23 bandpass filters in the first block. Figure taken from Moore and Glasberg, 1993.
2-17 Channels resulting from different amounts of recruitment on the vowel II taken from the word BIIDH. The vowel was at a level of 100 dB. Recruitment of $N = 1.0$ (where $N$ is the exponent to which the envelope is taken) corresponds to a hearing loss of 0 dB, recruitment of $N = 2.0$ corresponds to a flat hearing loss of 50 dB, and in general the recruitment amounts correspond to hearing losses via the equation $N = 100dB/(100dB - hearinglossdB)$. No additional offsets were used in creating the plot: the effective level changes with increasing recruitment due to the increased absolute threshold.

2-18 Channels resulting from different amounts of recruitment on the vowel II taken from the word BIIDH. The vowel was at a level of 80 dB. Recruitment of $N = 1.0$ (where $N$ is the exponent to which the envelope is taken) corresponds to a hearing loss of 0 dB, recruitment of $N = 2.0$ corresponds to a flat hearing loss of 50 dB, and the other recruitment amounts correspond to hearing losses via the equation $N = 100dB/(100dB - hearinglossdB)$. No additional offsets were used in creating the plot: the effective level changes with increasing recruitment due to the increased absolute threshold.
Channels resulting from different amounts of recruitment on the vowel II taken from the word BIIDH. The solid lines are the result of recruitment with a simulated loss of 0 dB, and the dotted lines are the result of recruitment with a flat loss of 50 dB. The dashed line shows channels made from the original signal without recruitment, showing that simulating a loss of 0 dB does not distort the signal very much. Signals with levels of 120, 100, and 80 dB were recruited corresponding to losses of 0 dB and 50 dB, creating pairs of lines for each signal level. The 120 dB signal is above 100 dB the limit of recruitment, so recruitment has little effect. The 100 dB signal has the highest parts at around the limit of recruitment, so they are not affected, while the higher frequency portions of the signal are lower in level and so are distorted. The 80 dB signal is severely distorted and reduced in level by recruitment.

Block diagram of $d'$ discrimination model. The bottom block comes after the top block in the processing path. The top block is used to create channels from all vowels processed, and the bottom block takes all combinations of pairs of vowels, and produces a $d'$ distance between each pair. In the top figure, the signal is first filtered into bands. Then, in each band, the average power per unit length of the signal is found by first squaring it, summing over the signal length, dividing by the signal length, and then normalizing to account for variation in signal length between stimuli. Finally, the natural log of the resulting number is taken to produce a channel. In the bottom diagram, the difference ($\Delta_i$) between corresponding channels is found for a pair of stimuli, as indicated in the plot. These differences are then put through the formula in the bottom of the figure. $\Sigma$ is an arbitrary constant that is a scale factor for all $d'$ distances.
2-21 Illustration of decision model function in two dimensions. Stimulus centers are represented by the small points in the figure, and are surrounded by Gaussian distributions in perceptual space. Response centers are found by taking the centroid of the stimulus centers of a single type. The probability distributions around each stimulus center are assumed to evoke responses corresponding to the closest response center. The lines dividing the figure are the boundaries of these so-called response regions.

2-22 Block diagram of decision model.

3-1 Total error rates for the simulation using subject JB's audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject JB's actual error rate.

3-2 Total error rates for the simulation using subject JO's audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject JO's actual error rate.

3-3 Total error rates for the simulation using subject MG's audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject MG's actual error rate.

3-4 Total error rates for the simulation using subject PW's audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject PW's actual error rate.
3-5  Total error rates for the simulation using subject RC’s audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject RC’s actual error rate. ................................. 98

3-6  Error rates relative to the total error rate for unsmeared stimuli for the simulation using subject JB’s audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject JB’s actual error rate. ....................... 99

3-7  Error rates relative to the total error rate for unsmeared stimuli for the simulation using subject JO’s audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject JO’s actual error rate. ....................... 100

3-8  Error rates relative to the total error rate for unsmeared stimuli for the simulation using subject MG’s audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject MG’s actual error rate. ....................... 101

3-9  Error rates relative to the total error rate for unsmeared stimuli for the simulation using subject PW’s audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject PW’s actual error rate. ....................... 102
3-10 Error rates relative to the total error rate for unsmear stimuli for the simulation using subject RC's audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject RC's actual error rate.

3-11 Plot showing how well the average of all of the confusion matrix data fit to the subject’s actual confusion matrices. Averages were taken across all subjects and smearing amounts, for the reduced-audibility simulation.

3-12 A comparison of the ter Keurs results (dotted line) and Moore results (solid line) if the total error rates are approximately the same for the two smearing conditions. The ter Keurs result had a smearing multiplier of 1.0, and the Moore smearing had a multiplier of 1.5. For each pair of ter Keurs and Moore values (plotted in the same vertical space), the subject used to create the simulated matrix and the real subject it was being compared to were the same.

3-13 A comparison of the ter Keurs results (dotted line) and Moore results (solid line) if the total error rates are approximately the same for the two smearing conditions. The ter Keurs result had a smearing multiplier of 1.5, and the Moore smearing had a multiplier of 2.0. For each pair of ter Keurs and Moore values (plotted in the same vertical space), the subject used to create the simulated matrix and the real subject it was being compared to were the same.
3-14 Graph showing the percentage of stimuli that had audible channels for a given channel number, for each subject. Note that the channels correspond to a log scale with channel 1 centered at 100 Hz, channel 12 at 1575 Hz, and channel 19 at 4839 Hz (for reference). Also, it can be seen that in most plots two groups of curves can be seen. The set of curves to the left corresponds to no recruitment, and the set to the right is with recruitment.

3-15 Plot showing the stimulus centers and response centers for the stimuli from the male and female speakers combined (top plot), and just the female speaker (bottom plot). The stimulus centers are shown by the small black dots, and the response centers are shown by the open circles. In each plot, the response centers are the average of the stimulus centers for each vowel. Only half of the total number of stimulus centers were plotted for computational reasons. As can be seen around the “AH” in the top plot, the stimuli corresponding to the male and female speakers formed two distinct clusters for each vowel. The clusters are not distinct for the other vowels in this plot due to the perspective. Also, it can be seen that the stimulus centers for each vowel are less spread out in the bottom plot.

3-16 Plot showing how well the average of all of the confusion matrix data fit to the subject’s actual confusion matrices. Averages were taken across all subjects, with the smearing multiplier fixed at 1.0, for the reduced-audibility simulation.

3-17 Total error rates for the simulation using subject JB’s audiogram, assuming that all frequencies are audible to the subject. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject JB’s actual error rate.
3-18 Total error rates for the simulation using subject JO's audiogram, assuming that all frequencies are audible to the subject. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject JO's actual error rate.

3-19 Total error rates for the simulation using subject MG's audiogram, assuming that all frequencies are audible to the subject. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject MG's actual error rate.

3-20 Total error rates for the simulation using subject PW's audiogram, assuming that all frequencies are audible to the subject. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject PW's actual error rate.

3-21 Total error rates for the simulation using subject RC's audiogram, assuming that all frequencies are audible to the subject. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject RC's actual error rate.

3-22 Plot showing how well the average of all of the confusion matrix data fit to the subject's actual confusion matrices. Averages were taken across all subjects and smearing amounts, for the simulation in which all frequencies were audible.
Assuming full audibility, a comparison of the ter Keurs results (dotted line) and Moore results (solid line) if the total error rates are approximately the same for the two smearing conditions. The ter Keurs result had a smearing multiplier of 1.0, and the Moore smearing had a multiplier of 1.5. For each pair of ter Keurs and Moore values (plotted in the same vertical space), the subject used to create the simulated matrix and the real subject it was being compared to were the same.

Assuming full audibility, a comparison of the ter Keurs results (dotted line) and Moore results (solid line) if the total error rates are approximately the same for the two smearing conditions. The ter Keurs and Moore results both had a smearing multiplier of 2.0. For each pair of ter Keurs and Moore values (plotted in the same vertical space), the subject used to create the simulated matrix and the real subject it was being compared to were the same.

Average vowel lengths for the different vowels. The error bars represent one standard deviation of the vowel length. The averages were nearly identical for male and female speakers.

Examples of spectra. In both cases, the $d'$ discrimination algorithm will produce equal $d'$ distances between all stimuli. However, a real human listener could easily distinguish that the third spectrum was much different than the other two. In Stimulus Set A, the spectrum heights are A, 2A, and A, for spectra 1, 2, and 3, respectively. The widths of the spectra are such that the total area under all of the spectra is the same.

Audiogram for subject PW.

Audiograms for subjects JB and JO.

Audiograms for subjects RC and MG.

Original simulation including reduced audibility.
<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-2</td>
<td>Original simulation including reduced audibility.</td>
<td>154</td>
</tr>
<tr>
<td>C-3</td>
<td>Original simulation including reduced audibility.</td>
<td>155</td>
</tr>
<tr>
<td>C-4</td>
<td>Original simulation including reduced audibility.</td>
<td>156</td>
</tr>
<tr>
<td>C-5</td>
<td>Original simulation including reduced audibility.</td>
<td>157</td>
</tr>
<tr>
<td>C-6</td>
<td>Original simulation including reduced audibility.</td>
<td>158</td>
</tr>
<tr>
<td>C-7</td>
<td>Original simulation including reduced audibility.</td>
<td>159</td>
</tr>
<tr>
<td>C-8</td>
<td>Original simulation including reduced audibility.</td>
<td>160</td>
</tr>
<tr>
<td>C-9</td>
<td>Original simulation including reduced audibility.</td>
<td>161</td>
</tr>
<tr>
<td>C-10</td>
<td>Original simulation including reduced audibility.</td>
<td>162</td>
</tr>
<tr>
<td>C-11</td>
<td>Original simulation including reduced audibility.</td>
<td>163</td>
</tr>
<tr>
<td>C-12</td>
<td>Original simulation including reduced audibility.</td>
<td>164</td>
</tr>
<tr>
<td>C-13</td>
<td>Original simulation including reduced audibility.</td>
<td>165</td>
</tr>
<tr>
<td>C-14</td>
<td>Original simulation including reduced audibility.</td>
<td>166</td>
</tr>
<tr>
<td>C-15</td>
<td>Original simulation including reduced audibility.</td>
<td>167</td>
</tr>
<tr>
<td>C-16</td>
<td>All frequencies audible condition.</td>
<td>168</td>
</tr>
<tr>
<td>C-17</td>
<td>All frequencies audible condition.</td>
<td>169</td>
</tr>
<tr>
<td>C-18</td>
<td>All frequencies audible condition.</td>
<td>170</td>
</tr>
<tr>
<td>C-19</td>
<td>All frequencies audible condition.</td>
<td>171</td>
</tr>
<tr>
<td>C-20</td>
<td>All frequencies audible condition.</td>
<td>172</td>
</tr>
<tr>
<td>C-21</td>
<td>All frequencies audible condition.</td>
<td>173</td>
</tr>
<tr>
<td>C-22</td>
<td>All frequencies audible condition.</td>
<td>174</td>
</tr>
<tr>
<td>C-23</td>
<td>All frequencies audible condition.</td>
<td>175</td>
</tr>
<tr>
<td>C-24</td>
<td>All frequencies audible condition.</td>
<td>176</td>
</tr>
<tr>
<td>C-25</td>
<td>All frequencies audible condition.</td>
<td>177</td>
</tr>
<tr>
<td>C-26</td>
<td>All frequencies audible condition.</td>
<td>178</td>
</tr>
<tr>
<td>C-27</td>
<td>All frequencies audible condition.</td>
<td>179</td>
</tr>
<tr>
<td>C-28</td>
<td>All frequencies audible condition.</td>
<td>180</td>
</tr>
<tr>
<td>C-29</td>
<td>All frequencies audible condition.</td>
<td>181</td>
</tr>
<tr>
<td>C-30</td>
<td>All frequencies audible condition.</td>
<td>182</td>
</tr>
</tbody>
</table>
D-1 Diagram of new hearing loss model. ......................... 188
List of Tables

2.1 Table of the levels at which the low-passed (below 2500 Hz) and high-passed (above 2500 Hz) portions of the CVC syllables were presented to the subjects in the confusion matrix experiment by Brantley (1990-1991). 41

2.2 Table of the different processing conditions used in the simulation. . . 71

3.1 Table of the average simulated confusion matrices for each processing condition. Averages were taken across subjects and smearing amounts. From top to bottom, the processing conditions are: No smearing, Moore smearing, ter Keurs smearing, and Hou smearing. The left column contains the results of smearing without recruitment, and the right column contains the results of smearing followed by recruitment. 83

3.2 Table of the simulated confusion matrices for approximately equivalent total smearing amounts for the Moore and ter Keurs methods, without recruitment. The results shown in the table are the averages across all subjects. The top row contains a Moore smearing with a widening factor of 4.5, and ter Keurs smearing of 0.75 octaves. The total error rate with that amount of smearing for the Moore method was a little lower than the total error rate for that amount of ter Keurs smearing. The bottom row contains a Moore smearing with a widening factor of 6, and ter Keurs smearing of 1.125 octaves. The total error rate with that amount of smearing for the Moore method was a little higher than the total error rate for that amount of ter Keurs smearing. . . . . . 85
3.3 Table of the average simulated confusion matrices for each processing condition. Averages were taken across subjects for a smearing multiplier of 1. From top to bottom, the processing conditions are: No smearing, Moore smearing, ter Keurs smearing, and Hou smearing. The left column contains the results of smearing without recruitment, and the right column contains the results of smearing followed by recruitment. ................................. 91

3.4 Table of the average simulated confusion matrices for each processing condition, assuming that all frequencies are audible. Averages were taken across subjects and smearing amounts. From top to bottom, the processing conditions are: No smearing, Moore smearing, ter Keurs smearing, and Hou smearing. The left column contains the results of smearing without recruitment, and the right column contains the results of smearing followed by recruitment. ................................. 92

3.5 Table of the simulated confusion matrices for approximately equivalent total smearing amounts for the Moore and ter Keurs methods, without recruitment, assuming that all frequencies are audible. The results shown in the table are the averages across all subjects. The top row contains a Moore smearing with a widening factor of 4.5, and ter Keurs smearing of 0.75 octaves. The total error rate with that amount of smearing for the Moore method was a somewhat (~5%) lower than the total error rate for that amount of ter Keurs smearing. The bottom row contains a Moore smearing with a widening factor of 6, and ter Keurs smearing of 1.5 octaves. The total error rate with that amount of smearing for the Moore method was almost identical to the total error rate for that amount of ter Keurs smearing. ................................. 93

A.1 Confusion matrix for subject JB. ................................. 138
A.2 Confusion matrix for subject JO. ................................. 138
A.3 Confusion matrix for subject MG. ................................. 138
A.4 Confusion matrix for subject PW. ............................... 139
A.5 Confusion matrix for subject RC. ............................... 139

B.1 Confusion matrices for the simulation done with subject JB’s parameters. From top to bottom, for each row the smearing conditions were: No smearing, Moore sm. only, ter Keurs sm. only, Hou sm. only. The amount of smearing in each column was the “Smearing Multiplier” times a ‘standard amount of smearing’ (see beginning of the appendix). 141

B.2 Confusion matrices for the simulation done with subject JB’s parameters. From top to bottom, for each row the smearing conditions were: Moore sm. then recruitment, ter Keurs sm. then recruitment, Hou sm. then recruitment, and just recruitment. The amount of smearing in each column was the “Smearing Multiplier” times a ‘standard amount of smearing’ (see beginning of the appendix). . . . 142

B.3 Confusion matrices for the simulation done with subject JO’s parameters. From top to bottom, for each row the smearing conditions were: No smearing, Moore sm. only, ter Keurs sm. only, Hou sm. only. The amount of smearing in each column was the “Smearing Multiplier” times a ‘standard amount of smearing’ (see beginning of the appendix). 143

B.4 Confusion matrices for the simulation done with subject JO’s parameters. From top to bottom, for each row the smearing conditions were: Moore sm. then recruitment, ter Keurs sm. then recruitment, Hou sm. then recruitment, and just recruitment. The amount of smearing in each column was the “Smearing Multiplier” times a ‘standard amount of smearing’ (see beginning of the appendix). . . . 144

B.5 Confusion matrices for the simulation done with subject MG’s parameters. From top to bottom, for each row the smearing conditions were: No smearing, Moore sm. only, ter Keurs sm. only, Hou sm. only. The amount of smearing in each column was the “Smearing Multiplier” times a ‘standard amount of smearing’ (see beginning of the appendix). 145
B.6 Confusion matrices for the simulation done with subject MG’s parameters. From top to bottom, for each row the smearing conditions were: Moore sm. then recruitment, ter Keurs sm. then recruitment, Hou sm. then recruitment, and just recruitment. The amount of smearing in each column was the “Smearing Multplier” times a ’standard amount of smearing’ (see beginning of the appendix).

B.7 Confusion matrices for the simulation done with subject RC’s parameters. From top to bottom, for each row the smearing conditions were: No smearing, Moore sm. only, ter Keurs sm. only, Hou sm. only. The amount of smearing in each column was the “Smearing Multplier” times a ’standard amount of smearing’ (see beginning of the appendix).

B.8 Confusion matrices for the simulation done with subject RC’s parameters. From top to bottom, for each row the smearing conditions were: Moore sm. then recruitment, ter Keurs sm. then recruitment, Hou sm. then recruitment, and just recruitment. The amount of smearing in each column was the “Smearing Multplier” times a ’standard amount of smearing’ (see beginning of the appendix).

B.9 Confusion matrices for the simulation done with subject PW’s parameters. From top to bottom, for each row the smearing conditions were: No smearing, Moore sm. only, ter Keurs sm. only, Hou sm. only. The amount of smearing in each column was the “Smearing Multplier” times a ’standard amount of smearing’ (see beginning of the appendix).

B.10 Confusion matrices for the simulation done with subject PW’s parameters. From top to bottom, for each row the smearing conditions were: Moore sm. then recruitment, ter Keurs sm. then recruitment, Hou sm. then recruitment, and just recruitment. The amount of smearing in each column was the “Smearing Multplier” times a ’standard amount of smearing’ (see beginning of the appendix).
Chapter 1

Introduction

1.1 Background

Due to the current state of technology, it is possible to do substantial signal processing in hearing aids. As such, the goal of hearing aid research is to come up with signal processing schemes to enable hearing-impaired individuals to experience sound as much as is possible like normal listeners, or at least to be able to consistently understand speech in various environments and at various levels. In order to develop said signal processing algorithms, it is necessary to understand how hearing-impaired people experience sound, so that any differences can be accounted for. Thus, hearing aid research currently is largely focused on developing models of how hearing-impaired people experience sound. This chapter contains a discussion of the physiology of hearing impairment and the perceptual consequences of hearing loss to provide background for understanding the models, and also discusses previous work on hearing loss simulations.

1.2 Physiology of Hearing Impairment

The first stages of the auditory system include the outer, middle, and inner ear, after which the auditory nerve takes the resulting signals to the brain. The outer and
middle ear serve to collect and channel the sound into the head. The inner ear serves several functions: it acts as a sort of frequency analyzer, to separate the sound into different frequency bands; it amplifies the sound; and it converts the mechanical sound signal into neural signals. These functions take place within the inner ear structure known as the cochlea. The cochlea is essentially a coiled tube that is partitioned lengthwise by several membranes. One of these is known as the basilar membrane (BM), which is a stiff structure that performs the first stage of frequency analysis through its mechanical properties: it is about 100 times stiffer at the base (the place where the sound first encounters it) than at the apex (the other end from the base). This variable stiffness means that different locations on the BM will respond (oscillate) the best to different input frequencies: it acts as a frequency-to-place converter. The frequency that causes the maximum response of a particular location on the BM is known as the best frequency (BF) or characteristic frequency (CF) of that location. High frequencies cause the basal end to be excited more, while low frequencies travel down the length of the BM and create more of a response toward the apical end.

Attached to the BM are rows of cells known as outer hair cells (OHCs). These cells serve primarily as mechanical transducers, actively causing the BM to vibrate even more in response to incoming sounds than it would on its own. This causes the BM to be even more frequency selective (that is, exhibit sharper tuning to tone stimuli) and causes the incoming sounds to be amplified. It has been found that due to the OHCs, the transfer function between the incoming sound level and the resulting BM oscillation displacement at a particular location is compressive for CF tones. This compression takes place for tones at input levels between about 30-90 dB SPL (Oxenham and Plack, 1997). A graph showing this transfer function is shown in figure 1-1. For tones that are an octave or more lower than the CF of a location on the BM, the BM responds linearly. This is also shown in figure 1-1.

The motion of the BM is detected by another type of cell that is found along the length of the BM, the inner hair cell (IHC). These cells connect to auditory nerve
Figure 1-1: Figure showing how the basilar membrane is compressive for tones at the best frequency of the location of measurement. This figure shows "the level of a masker required to mask [a] 6-kHz [tone] signal, as a function of signal level" (Oxenham and Plack, 1997). The plot of the on-frequency masker is linear, as would be expected. However, the off-frequency masker shows compression, indicating both that the basilar membrane is compressive and that it responds linearly for tones an octave below the best frequency of a given location. Taken from Oxenham and Plack, 1997.
fibers, and act to convert the amplitude (or velocity) of BM oscillation into neural firing patterns. In response to frequencies below about 2 kHz, the IHCs cause spikes on the auditory nerve to be "phase-locked" to those frequencies, meaning that the spike trains are in phase with the BM motion caused by the signal, and the temporal patterns of the signal itself. At higher frequencies, auditory nerve fiber firing patterns are not phase-locked to the BM motion, meaning that their firing rate is not correlated with the signal.

When sensorineural hearing loss occurs, typically the OHCs and/or IHCs are damaged. This damage can occur in any location along the length of the cochlea, which translates to different frequency regions having different amounts of hearing loss due to the frequency-to-place transformation in the cochlea. IHC loss is primarily a conduction loss, as the IHCs act as sensors of the BM motion. This type of hearing loss can typically be corrected for by linear amplification of the incoming signal. However, in the majority of cases, OHC loss is the more dominant effect in hearing loss. The effect of OHC loss is that the sharp frequency tuning of the BM is lost in regions where there has been OHC damage, since the OHCs will no longer function as active amplifiers. Additionally, those regions of the cochlea do not have the gain in signal level caused by the OHCs. This means that compression will not take place to as large a degree. In cases of severe OHC loss, no compression or signal gain remains, and the BM responds linearly to the incoming sound waves.

The current experiment, and hearing aid research in general, are focused on studying the effects of hearing loss caused by OHC damage alone. The next section will discuss the perceptual consequences of this type of hearing loss.

1.3 Psychoacoustics of Hearing Impairment

In discussing the perceptual consequences of hearing loss, it will be helpful to first define some terms commonly used in discussing auditory perception. The first idea is the "auditory filter." If the transfer function from a sound in air to a person's internal
perception of the sound is examined, a good model for this transfer function is a filter bank of auditory filters. The specific form of the auditory filters will be described later, but in general their center frequencies are logarithmically spaced above around 600 Hz (below this they have constant spacing). For moderate input sound levels the filters are symmetric, but for high input sound levels they are asymmetric and respond more to frequencies higher than the center frequency. The next idea is the “excitation pattern,” which is defined as “the output from the auditory filters as a function of filter center frequency...” (Baer and Moore, 1993). Or in other words, if a sound is played to a listener, the excitation pattern is the plot of how much each auditory filter responds versus the center frequencies of the auditory filters.

The psychoacoustic effects of OHC loss are generally described in terms of loudness recruitment and absolute threshold reduction, and frequency and temporal smearing. These will be discussed in turn.

Loudness recruitment is a phenomenon that stems from the reduced gain of the cochlea in regions where there is OHC loss. In normal ears, quiet sounds are amplified to a moderate level through the compressive behavior of the cochlea. However, at and above about 100 dB, the incoming sounds are no longer amplified (Moore et al., 1996). In impaired ears, quiet sounds are not amplified, and so below a certain level threshold, sounds cannot be heard at all. This is an absolute threshold elevation. For sounds just above the threshold, the perceived loudness increases more rapidly with increases in signal level than it would for normal-hearing listeners. This is because of the lack of or reduced compression in the cochlea, and because the perceived loudness is related to the amount of BM motion. A graph showing equal-loudness contours for subjects with unilateral hearing losses is shown in figure 1-2.

Frequency and temporal smearing occurs because of the less-sharp tuning of the BM in an impaired cochlea. Locations on the BM at which OHC damage has occurred respond to a wider range of frequencies than they would in a normal ear. The psychoacoustic effect of this is that an impaired-listener’s auditory filters are wider than they are supposed to be, and that the excitation patterns produced also less
Figure 1-2: Loudness matching curves for subjects with unilateral hearing loss. The thin line at a 45-degree angle is the expected result if the two ears gave equal loudness results for a tone at a given level. The fact that the actual data is at a steeper slope than this indicates that for the impaired ear, as the stimulus sound pressure level increases a little, the apparent loudness increases more rapidly than in a normal ear. Also, the asterisks indicate the subjects' absolute thresholds; these are elevated in the impaired ear relative to the normal ear. Taken from Moore et al., 1996, with the 45-degree line added.
sharp. Also, since a sum of a lot of frequencies (which is what will come through the wider auditory filter) produces a less-regular temporal pattern than a single frequency or just a few frequencies (which is what would come through normal-width auditory filters), the temporal patterns at the output of each auditory filter are more complex.

Moore et al. (1992) provide a convenient summary of the psychoacoustic changes produced by hearing impairments:

(1) It would produce broader excitation patterns for sinusoidal signals. For complex signals containing many harmonics (e.g. speech sounds) this would make the individual components harder to resolve.

(2) For complex signals such as speech it would result in a blurring of the spectral envelope, making spectral features such as formants harder to detect and discriminate (Leek et al., 1987).

(3) It would change the time patterns at the outputs of the auditory filters, generally giving more complex stimulus waveforms (Rosen and Fourcin, 1986).

It is also important to note that these perceptual effects of hearing loss are all closely related: it is likely not possible to create neural stimuli that code for one effect but not the others (Moore et al., 1992). In the processes of changing either the frequency or temporal patterns in a signal, the perception of the other will almost certainly be changed due to their being coded in the same IHC responses.

1.4 Simulating Hearing Loss

1.4.1 Reasons for Simulating Hearing Loss

As stated in the background, in order to develop signal processing algorithms for use in hearing aids, it is necessary to understand how hearing-impaired people experience sound, so that any differences can be accounted for. It would be very useful to
have a computer model that would accurately represent hearing-impaired listeners' responses, because then simulations could be done using the model to predict real listeners' responses and provide the optimal hearing aid correction. Also, it is useful to model the psychoacoustic effects of hearing loss separately. Though it is not possible to isolate the effects in real listeners, it would be beneficial to determine which effects cause the most deleterious effects on speech reception, because hearing aid signal processing schemes may be able to correct for one effect (at a possible expense to the other effects, but with a net benefit to the hearing aid user). Finally, hearing loss models can be used to educate normal-hearing listeners about hearing impairments. Such education could be useful for people who interact with hearing-impaired listeners frequently.

1.4.2 Previous Work Simulating Hearing Loss

Simulations of hearing loss were first produced the 1930s, when Steinberg and Gardener (1937) used masking noise to simulate increased hearing thresholds and loudness recruitment. However, with masking noise the mechanisms that occur in the ear and brain to cause the percept of hearing loss in normal hearing listeners are thought to be different than the mechanisms of hearing impairment in real hearing-impaired listeners (Graf, 1997). More recently, advances in signal processing capabilities have allowed attempts to be made at more accurate hearing loss simulations. Generally, studies have attempted to generate stimuli that, when played to a normal-hearing listener, simulate one aspect of hearing loss at a time. Some studies have simulated the aspects of reduced time or frequency resolution, and some have simulated loudness recruitment. A few studies have used a recruitment model in conjunction with a reduced-resolution model to try simulate hearing loss more generally. However, as was pointed out in the previous section, it is not generally possible to create signals that, when presented to normal-hearing listeners, will evoke a completely accurate perception of hearing loss in all respects.

Moore and Glasberg (1993) simulated simulated threshold elevation and loudness
recruitment with a method used later in the current experiment. They simulated moderate and severe flat hearing loss conditions, and a sloping moderate-to-severe hearing loss condition. They examined responses of normal-hearing subjects listening to the stimuli, which were sentences containing key words. They found that with speech processed in quiet, at low presentation levels the intelligibility was reduced. If the speech was amplified to normal conversation levels before processing (simulating a linear hearing aid), then it was highly intelligible, however. In contrast, if the speech contained a competing talker in the background, intelligibility was compromised even at high presentation levels. These results were extended by Moore et al., 1995, who simulated recruitment in the same way except with a background of speech-shaped noise. They found that intelligibility was reduced for the unamplified case, but if linear amplification was applied then intelligibility was almost as good as in the unprocessed (control) condition—performance decreased much less than with the single talker interference in the previous experiment.

A number of experiments have simulated reduced frequency selectivity. Baer and Moore (1993) processed sentences and presented them at a moderate level to normal-hearing listeners, using an algorithm also used in the current simulation. They found that if speech was processed in a background of quiet, even simulating auditory filters with 6 times the normal width did not affect the intelligibility very much. If the speech was processed in a background of speech-shaped noise, intelligibility significantly decreased. The intelligibility decreased more with larger amounts of smearing and with lower signal-to-noise ratios. The authors state that these results are consistent with the behaviors of real hearing-impaired listeners. Another study by Baer and Moore (1994) extended these results by processing speech in the presence of a competing talker. They found that intelligibility was reduced, but not a much as with the noise background.

ter Keurs et al. (1992) tested a different spectral smearing algorithm, also used in the current simulation. They examined both the speech reception threshold (SRT) for sentences and in a second experiment, vowel (and consonant—not discussed here)
confusions caused by the smearing. They found that the subjects' SRT increased once the smearing bandwidth increased beyond the ear's critical bandwidth, about 1/3 to 1/2 octaves, and continued to increase after this point. In the confusion matrix experiments, the total error rate for vowel confusions remained close to zero until somewhere between 1/2 and 2 octaves of smearing. The particular confusions causing the errors tended to be with the back vowels. These results held for both vowels in quiet and in a noise background. A second experiment by ter Keurs, et al. (1993) examined SRTs to sentences spoken by a male speaker (the first experiment had used a female speaker) and found similar results regarding the critical bandwidth.

Hou and Pavlovic (1994) performed temporal smearing on vowel-consonant nonsense syllables to evaluate its effect on speech intelligibility, using an algorithm repeated in the current experiment. They found that with small smearing durations, speech was not degraded. Even with the larger durations, speech intelligibility was only reduced when the signals were low-pass filtered. Also, they demonstrated that the spectral characteristics of the signals were largely unaltered.

Nejime and Moore (1997) simulated the combined effects of frequency smearing followed by loudness recruitment on speech intelligibility, using the models from Baer and Moore (1993) and Moore and Glasberg (1993). Their stimuli were sentences in speech-shaped noise. They tested a simulated moderate flat loss or moderate-to-severe sloping loss for both frequency smearing and loudness recruitment, and found that speech intelligibility was degraded substantially, and could not be corrected for by amplification alone.

In general, these studies have shown that simulations of frequency smearing and loudness recruitment degrade speech intelligibility for sentences with key words, for both male and female speakers. Temporal smearing does not seem to have as much of an effect on speech intelligibility. However, aside from the ter Keurs et al. (1992) experiment, these studies have not attempted to verify the accuracy of the smearing methods aside from gross intelligibility rates. Even in the ter Keurs (1992) paper, a comparison between the confusion matrix resulting from simulated smearing and a
confusion matrix corresponding to real hearing-impaired listeners was not provided. So, questions remain about how well these smearing methods simulate real hearing impairments.

1.5 Problem Statement

Most previous studies on hearing loss simulation have examined the amount of speech intelligibility degradation caused by the hearing loss simulations without examining how well the percept evoked in normal-hearing listeners matched what hearing-impaired listeners experience. The simulations described in this thesis attempt to evaluate different hearing loss models of reduced frequency or temporal resolution and loudness recruitment/threshold elevation in order to determine their accuracy. To accomplish this task, a model of human speech perception is employed. This provides a deterministic way of evaluating the hearing loss simulations, and allows for multiple parameter values to be considered efficiently. It is also convenient that the stimuli under consideration, vowels, are determined largely by their spectra; the perceptual model being used operates solely on the basis of signals' spectra, allowing discrimination to proceed without interference from other signal characteristics.
Chapter 2

Signal Processing in the Experiment

2.1 Overview of Experiment

This thesis involved simulating hearing loss using several models. In particular, the models of hearing loss that were being compared and evaluated performed frequency smearing and loudness recruitment. These models converted waveforms corresponding to ordinary speech into waveforms that would evoke the percept of hearing loss in normal-hearing listeners.

It was desired to determine how well these models simulated what real hearing-impaired people experience when they hear sounds. In order to determine what people experience, something was needed to convert the physical sound into a person’s perception of the sound. This could either be a real person (i.e., subjects could listen to sounds processed by the hearing loss models and report on their perceptions), or it could be a model simulating how normal-hearing people perceive sounds. In this thesis, the latter option was decided upon in order to try accurately simulate the entire pathway from a sound in air to what a hearing-impaired listener would experience in response to that sound. The combination of a hearing loss model followed by a perceptual model would ideally react the same way to stimuli as would a real person.
So, to test this simulation, the same stimuli were presented to real hearing-impaired subjects as well as to the simulation, and the responses of each were compared.

2.2 Overview of Signal Processing

A block diagram showing an overview of all the signal processing steps that were done is shown in figure 2-1. The bottom path in the block diagram shows the processing by real subjects, while the top path shows the steps in the simulation. The stimuli used to compare the simulation to real subjects were a set of CVC tokens, approximately half spoken by a male speaker, and about half spoken by a female speaker.

2.2.1 Processing by Real Subjects

The real hearing-impaired subjects performed their “processing” as follows: five hearing-impaired subjects listened to the set of CVC tokens, and confusion matrices were generated based on the subject’s responses. This step was done by Merry Brantley in the Sensory Communication group of the Research Laboratory of Electronics at MIT in 1990 and 1991. Although confusion matrices for both the vowels and the consonants were generated in the previous experiment by Brantley, only the vowel confusion matrices were used in the current research.

2.2.2 Processing by the Simulation

In the simulation, processing was done as follows: first, the vowels were extracted from the CVC tokens. Next, the extracted vowels were put through the hearing loss models, which involved simulating frequency smearing, loudness recruitment, or a combination of the two. The models used in this section were taken from papers by Baer and Moore (1993), ter Keurs et al. (1992), Hou and Pavlovic (1994), Moore and Glasberg (1993), Farrar et al. (1987), and Braida (1991). The hearing loss models produced a set of waveforms that would sound to normal-hearing listeners as if the listeners had hearing loss. The combination of the vowel extraction block and the
Figure 2.1: Signal processing overview.
simulated hearing loss block formed the hearing loss model portion of the simulation. The subsequent stages in the simulation made up the perceptual model portion of the simulation.

The first stage of the perceptual model shall be referred to as the "d' discrimination algorithm," and it was concerned with determining how different two sounds would be perceptually. This model was created by Farrar, et al. (1987). The first stage of the model, which shall be referred to as the "make channels" block, took as an input a single vowel waveform and essentially found the signal power in a number of consecutive frequency bands. The outputs from this model will be referred to as "channels." The next stage of the d' discrimination algorithm took the channels from two different vowels, and computed a d' distance between them. This d' distance was interpreted as being the distance between the two vowels in "perceptual space," that is, how different they would be perceived to be by a human listener. In response to the set of processed vowel waveforms, the d' discrimination algorithm generated a matrix of d' distances between all vowel pairs.

The second stage of the perceptual model converted this set of d' distances into a confusion matrix. This was done by the decision model block. This block first used classical multi-dimensional scaling to produce a map of the vowels that was reflective of the d' between-vowel distances. This map was also interpreted as being in perceptual space, as its dimensions were in units of d'. To take into account of the perceptual uncertainty that is present if a person is exposed to a stimulus, each point was converted into a Gaussian probability distribution centered at the point's location. Finally, a decision algorithm was applied as follows: the centroid of the locations for each vowel was computed. These centroids were interpreted to be the "response centers" in perceptual space, while the individual vowel locations surrounded by Gaussians were interpreted to be the "stimulus centers." An ideal detector algorithm was used such that the stimulus centers (or really, the parts of the Gaussians around them) closest to a response center were classified as corresponding to that response. In this manner a confusion matrix was generated corresponding to
the processed sounds, completing the simulation pathway.

Finally, to compare the responses created by the simulation to those generated by real listeners, a Chi-squared-like measure was used to evaluate how closely the confusion matrices matched.

2.3 Detail on the Brantley Confusion Matrix Experiment

The experiment that produced the confusion matrices for this thesis was examining the effect of filtering on audibility in hearing-impaired listeners, and was done by Brantley in 1990-1991. The five subjects in this experiment had moderate or moderate-severe hearing losses; three subjects’ losses were flat, and two were sloping. Audiograms for the five subjects are presented in figures A-1, A-2, and A-3, in Appendix A.

The experiment used part of a set of 862 consonant-vowel-consonant (CVC) syllables of the form /æ/-CVC, where /æ/ is the unstressed schwa. The set of CVC syllables had the vowels /i, a, u, i, e, u/, and the 16 consonants /p, t, k, b, d, g, f, θ, s, ʃ, v, ð, z, c, f, h, z/. The Brantley experiment limited their stimuli to the 12 consonants /p, t, k, b, d, g, s, ʃ, θ, v, z, ð/, so they only used about 484 syllables. The mean durations of the syllables spoken by the male speaker were 634 ms, and the syllables spoken by the female speaker had a mean duration of 574 ms (the vowels used in this thesis spoken by the male and female speakers had approximately equal durations). The stimuli in the Brantley experiment were lowpass filtered at 9 kHz, and digitized at a sampling rate of 20 kHz with 12-bit resolution.

Each stimulus was filtered into a low-passed portion (below 2500 Hz) and a high-passed portion (above 2500 Hz). Each portion was presented separately to the five subjects in a background of quiet, and the level was adjusted to maximize the comprehension of the CVC stimuli. Then, the low-passed and high-passed portions of the signal were added together, using the levels found in the first part of the
<table>
<thead>
<tr>
<th>Subject</th>
<th>Low-passed portion [dB SPL]</th>
<th>High-passed portion [dB SPL]</th>
</tr>
</thead>
<tbody>
<tr>
<td>JB</td>
<td>80</td>
<td>95</td>
</tr>
<tr>
<td>JO</td>
<td>110</td>
<td>105</td>
</tr>
<tr>
<td>MG</td>
<td>107</td>
<td>110</td>
</tr>
<tr>
<td>PW</td>
<td>85</td>
<td>88</td>
</tr>
<tr>
<td>RC</td>
<td>80</td>
<td>102</td>
</tr>
</tbody>
</table>

Table 2.1: Table of the levels at which the low-passed (below 2500 Hz) and high-passed (above 2500 Hz) portions of the CVC syllables were presented to the subjects in the confusion matrix experiment by Brantley (1990-1991).

experiment, and a confusion matrix experiment was done. In the experiment, each stimulus was presented to each subject 50 times, and confusion matrices for the vowels and consonants were generated. A table of the levels at which the high-passed and low-passed portions of the signals were presented to each subject is shown in table 2.3.

Since the simulations done in this thesis attempted to duplicate this confusion matrix experiment, a simulation was done using the parameters for each subject (hearing loss and levels used in the experiment). It was hoped that the results of the simulations would correspond to the confusion matrices produced by the respective subjects. The description of the simulation in the following sections describes it as if just one set of parameters was being used, but in reality the simulation was run five times, corresponding to once for each subject. The confusion matrices produced by the real subjects are shown in Appendix A.

### 2.4 Speech Stimuli and Vowel Extraction

The stimuli used in the simulation were the same as described above in the section on the Brantley confusion matrix experiment, except that the simulation used the entire set of stimuli with 16 possible consonants, instead of being limited to 12 consonants as in the actual experiment. Also, the sampling rate of the stimuli in this thesis was 10 kHz instead of 20 kHz, thereby limiting the bandwidth of the stimuli to 5 kHz.
For the remainder of this thesis, the vowels /a, i, ɛ, ɪ, u, ʊ/ will be referred to as AH, EE, EH, II, OO, and UH, respectively.

Though these tokens contained both consonants and vowels, only the vowels were used in the simulation. The vowels were thus extracted from the entire word and used in isolation. This is another difference between the Brantley confusion matrix experiment and the simulation in this thesis: in the Brantley experiment, the vowels were heard in the context of consonants, but in this simulation they were in isolation. Vowel extraction was achieved by making use of the fact that the energy in the vowels was usually much greater than that of the consonants. The waveform for each word was squared, finding the power, and the envelope of this was found using a second-order lowpass filter with a cutoff frequency of about 10.3 Hz. Next, a threshold was set to be 0.2 times the maximum value of power envelope. The longest stretch of contiguous signal above this threshold was judged to be the vowel (this metric was correct in all cases to identify the general region of the vowel). Next, to eliminate the transition regions between the vowel and surrounding consonants, 20 ms of signal was removed from the beginning of the vowel, and 35 ms of signal was removed from the end of the vowel. The extracted vowels were listened to in order to verify that the vowels had been extracted well. A few vowels were removed from the set due to clicks or other signal artifacts. The syllable waveforms at this point had a sampling rate of Fs=10000. They were normalized to have a specified average power per unit time. It was also at this point that the signal was scaled so that the levels above and below 2500 Hz matched those that were used in the confusion matrix experiment with real subjects, for a given subject.

2.5 Signal processing simulating hearing loss

2.5.1 Methods of frequency smearing

There were three methods of frequency smearing considered; or more specifically, two methods simulated reduced frequency selectivity in the impaired ear (Baer and
Moore, 1993 and ter Keurs, et al., 1992), and one method simulated reduced temporal resolution in the impaired ear (Hou and Pavlovic, 1994). These methods shall be referred to as the “Moore,” “ter Keurs,” and “Hou” methods in short. In all of these methods, the purpose of processing was to generate a time waveform that would sound to normal-hearing listeners as if they had hearing loss.

Moore Smearing

The “Moore” method was first presented in Moore et al. (1992), then modified somewhat in Baer and Moore (1993) and later expanded upon and further modified by Nejime and Moore (1997). The method used in this experiment is that presented in the Baer and Moore (1993) paper. This smearing method attempted to create an excitation pattern for a normal-hearing listener evoked by a processed sound that would resemble the excitation pattern for an impaired-hearing listener by a normal sound. This attempted to duplicate the effect of the wider-than-normal auditory filters and reduced frequency selectivity found in impaired-hearing listeners. The particular degree of widening was intended to resemble that found in individuals with “moderate or severe cochlear hearing loss” (Baer and Moore, 1993). This processing notably did not attempt to duplicate any time-domain features of what hearing-impaired listeners would experience, but only spectral differences, as it is likely impossible to accomplish both simultaneously (Baer and Moore, 1993).

Block diagrams of the processing done in this method are shown in figure 2-2. This method of smearing used overlap-and-add to process the input signal over time. 8 msec chunks of the waveform were extracted, then windowed, and an FFT was done on each chunk. Next, the FFT was separated into the amplitude and phase components. The phase component was left unmodified during the processing. The amplitude component was squared to generate the power, and then was smeared as follows. Auditory filters of normal width and of a width N times larger than normal were calculated, with center frequencies at each of the frequencies in the FFT. These auditory filters were stored in a matrix, such that if a FFT was multiplied by the
matrix, each frequency in the FFT would contribute its own excitation pattern to the final result. In other words, the end result of an FFT multiplied by the matrix was the excitation pattern generated by a stimulus with that FFT, and using those auditory filters.

To create the effect of normal-hearing listeners having widened auditory filters, the signal was multiplied by the wider-than-normal filter matrix in order to generate the effect of wider auditory filter, and then multiplied by the inverse of the normal-width filter matrix. This procedure removed the effect of a normal-hearing listener’s own auditory filters, leaving just the wider-than-normal filters to create an excitation pattern. After this smearing procedure, the square root of the result was taken to convert back to amplitude, and this was combined with the unmodified phase. An inverse FFT was taken, and then different processed chunks of the signal were combined using the overlap-add procedure to create the final output waveform.
The particular details of the processing were as follows. The original vowel signals, at a sampling rate of Fs=10000 Hz, were resampled to 16000 Hz for the Moore processing. Next, the vowel waveforms were divided up into frames for processing (analogous to the inverse of overlap-and-add). Frames were of length 128 samples, corresponding to 8 ms, with a shift of 64 samples (4 ms) between frames. The frames were then windowed with Hamming windows of length 128 samples, then padded with 64 zeros on each side, to create a total length of 256 samples. An FFT of each frame was then taken, and separated into the amplitude and phase. The amplitude component was squared, then multiplied by two matrices. One represented auditory filters widened by a factor of \( N \), and one was the inverse of a matrix of normal-width auditory filters. The auditory filters were symmetric and were of the form of a roex\((p)\) filter:

\[
W(g) = (1 + pg)\exp(-pg),
\]

“where \( g \) is the [frequency] deviation from the center frequency \( f_c \) of the filter divided by \( f_c \).” \( W(g) \) is the resulting filter shape, and \( p \) is a measure of the sharpness of the filter. \( p \) is calculated using the following equations:

\[
(2.2) \quad ERB = 24.7(0.00437 f_c + 1).
\]

\[
(2.3) \quad p = 4f_c/ERB,
\]

In these equations (and elsewhere), \( ERB \) is the equivalent rectangular bandwidth of the filter. To simulate widening of the auditory filters, the value of \( p \) calculated via these equations was then divided by the widening factor \( (N) \). Or equivalently, the calculated \( ERB \) was multiplied by the widening factor before calculating \( p \). For the normal-width filters, \( p \) was not modified. The frequencies in the FFT (and hence the frequencies at which the auditory filters were centered on) were from 0 to 8000 Hz every 62.5 Hz.

After the smearing was performed, the frequencies higher than 6875 Hz (the 110th
frequency) were set to 0. Next, the square root of the result was taken to generate the spectrum amplitude instead of power. Finally, the original phase was recombined with the modified spectrum, and an inverse FFT was taken to create a time-domain signal of length 256. These 256-point frames were recombined using the overlap-add method with a shift between frames of 64 samples. In the Baer and Moore paper (1993), the resulting signal was low-pass filtered at 7 kHz both before and after the smearing process; however, this was not done in the current implementation of the smearing algorithm. Finally, the power of the smeared waveform was normalized to equal the power of the original input signal, in order to keep the level constant across stimuli. Examples of the outputs of the processing by the Moore method in the present simulation are shown in figures 2-3 and 2-4. The second figure shows the spectrum converted into “channels,” which essentially shows the excitation pattern produced by the signal. The process of creating these channels will be explained later.

It should be noted that in the implementation in this research, symmetric auditory filters were used. This assumption is realistic at moderate sound levels, but both the final stages of processing and the real hearing-impaired listener experiment used high sound levels, such that this assumption would not be accurate. This is one potential source of error in the results, and will be discussed later.

Also, the Baer and Moore (1993) paper discussed the actual effective degree of smearing created by the model relative to the intended degree of smearing. This is illustrated in figure 2-5, showing examples of a vowel spectrum before and after smearing. They found that as a result of the overlap-add procedure, the final waveform had an effective smearing degree of much less than specified. For an intended smearing factor of 3, a representative vowel had an actual smearing factor of about 1.5, as determined by a least-squares matching over the formant-containing frequencies. This result also applies to the present simulation, and it is likely that it applies to the ter Keurs, et al. (1992) algorithm as well, which also uses the overlap-add method. The ter Keurs, et al. algorithm used longer windows than the Baer and Moore algorithm, so the overlap-add method may have less of an effect on reducing
Figure 2-3: Spectra resulting from different amounts of Moore smearing on the vowel II taken from the word BIIDH. All files were put through the recruitment algorithm but with no recruitment. The unsmeared signal is shown at the regular level, while the other traces have been offset by -40 dB increments. The labels to the left of the traces indicate the amount of smearing, in terms of the effective number of ERBs of the widened auditory filters; US indicates the unsmeared signal.
Figure 2-4: Channels resulting from different amounts of Moore smearing on the vowel II taken from the word BIIDH. All files were put through the recruitment algorithm but with no recruitment. The unsmeared signal is shown at the regular level, while the other traces have been offset by -8 dB increments. The labels to the left of the traces indicate the amount of smearing, in terms of the effective number of ERBs of the widened auditory filters; US indicates the unsmeared signal.
the effective degree of smearing.

**ter Keurs**

Like the Moore method of smearing, the ter Keurs method (ter Keurs et al., 1992, and ter Keurs et al., 1993) simulated only reduced frequency resolution. Unlike the Moore method, the algorithm for smearing did not attempt to simulate the frequency resolution of hearing-impaired listeners exactly; rather, the algorithm was a more general-purpose smearing algorithm that simulated reduced frequency resolution. The model did not smear the actual spectrum of the signal, as did the Moore model, but rather the spectral envelope, and then imposed the smeared spectrum on the original signal: ter Keurs et al. (1992) write, “in this way spectral contrasts [are] gradually reduced, but the signal’s phase and harmonic structure [are] preserved.” The smearing was accomplished by first projecting the spectral envelope onto a log-frequency scale, then convolving it with a Gaussian-shaped filter that effectively acted as a low-pass filter. The convolution acted to average the envelope over \( B \) octaves, where \( B \) is the effective ERB of the filter.

The overall processing system used in the ter Keurs et al. (1992) model is shown in figure 2-6, and the specifics of the spectral smearing are shown in more detail in figure 2-7. A sample of the output of the model used by ter Keurs et al. is shown in figure 2-8, while samples of the output of the implementation of the ter Keurs method used in the current simulation (which differs from the original slightly) are shown in figures 2-9 and 2-10. The signal processing sequence is rather similar to the Moore method, in general; it divides the signal into frames, does an FFT, smearing, an inverse FFT, then creates the new signal via the overlap-add method. In detail, the processing was as follows. The vowel waveforms were first resampled to be at \( Fs=15625 \), having previously been at \( Fs=10000 \). (The waveforms used in the experiments run by ter Keurs et al. (1992 and 1993) were bandlimited to 6250 Hz.) A vowel to be smeared was then divided into 256-sample frames (corresponding to 16.4ms) with a shift of 64 samples (4.1ms) between frames. Each frame was then windowed using a Hamming
Figure 2-5: Sample spectra showing the results of smearing, for a synthetic vowel /ae/. The different frames (vertical dimension in the figure) show successive chunks of signal that are partially overlapping. As can be seen in the right column, showing the output spectra, the effective degree of smearing is less than was intended. The intended degree of smearing is seen in the middle column, which shows the results after smearing but before the overlap-add process. Taken from Baer and Moore, 1993.
window, and padded with 384 zeros on each side to make a total length of 1024 samples. An FFT was taken of the result, and the power spectrum was found by squaring the amplitude of the FFT.

At this point, the processing done in the current research differed slightly from that which was described in the ter Keurs et al. papers. In the current simulation, the envelope of the power spectrum was found by low-pass filtering it twice (in succession) with a first-order low-pass filter with a 3dB frequency of 894.4 Hz. In the paper, the spectral envelope was found by taking the log of the power spectrum then an inverse FFT to find the signal’s cepstrum. In ter Keurs et al. (1992), they write about the subsequent processing, “This cepstrum is “liftered” and through a forward FFT the (log) spectral envelope is obtained. The envelope is then adjusted to run smoothly over the spectral peaks.” It is unclear whether or not in ter Keurs et al.’s implementation at what point the inverse log of the spectral envelope was taken—they might have done the convolution with the Gaussian filter with the log spectral envelope rather than the linear spectral envelope, as was done in the current simulation. This is another potential way that the processing techniques could have differed.

After the spectral envelope was found, it was projected onto a log base 2 frequency scale. In the current implementation, since this projection is non-uniform, intermediate values were interpolated in a stepwise manner—the values between one known point and the next were set to value of the lower frequency point. Next, the log-scale power spectrum was smeared by convolving it with a Gaussian of the following form:

\[
W(\log_2 f) = e^{-\pi (\log_2 f/B)^2},
\]

where \( \log_2 f \) is the base 2 logarithm of the frequency scale, and \( B \) is the ERB of the Gaussian, in octaves, as previously stated. Further details about the mathematics behind this smearing process can be found in ter Keurs et al. (1993). After the log-scale power spectrum was smeared, it was converted back to a linear frequency
Figure 2-6: Block diagram showing the details of the ter Keurs processing method. Taken from ter Keurs et al., 1992.
Figure 2-7: Block diagram showing the details of the Spectral Smearing block from 2-6. Taken from ter Keurs et al., 1992.
Figure 2-8: Sample of the outputs of the ter Keurs frequency smearing algorithm. The top log spectrum, labeled “US,” shows the unsmeared signal, while the bottom three spectra show the results of smearing of $1/8$, $1/2$, and $2$ octaves, as labeled. Taken from ter Keurs et al., 1992.
Figure 2-9: Spectra resulting from different amounts of ter Keurs smearing on the vowel II taken from the word BIIDH. All files were put through the recruitment algorithm but with no recruitment. The unsmeared signal is shown at the regular level, while the other traces have been offset by -50 dB increments. The labels to the left of the traces indicate the amount of smearing, in octaves; US indicates the unsmeared signal. One octave of smearing is approximately equivalent to 3.9 ERBs of smearing.
Channels corresponding to signals with different amounts of ter Keurs smearing.

Figure 2-10: Channels resulting from different amounts of ter Keurs smearing on the vowel II taken from the word BIIDH. All files were put through the recruitment algorithm but with no recruitment. The unsmeared signal is shown at the regular level, while the other traces have been offset by -12 dB increments. The labels to the left of the traces indicate the amount of smearing, in octaves; US indicates the unsmeared signal. One octave of smearing is approximately equivalent to 3.9 ERBs of smearing.
scale, then imposed on the original spectrum by multiplying the original by the ratio of the old and new amplitudes (i.e., the square root of the power spectrum). The inverse FFT was taken, and the result was then windowed again. This second window was of length 1024 samples, and had a “unity central part of 256 samples and sine wave-squared skirts.” Finally, the output signal was reconstructed by overlap-adding the frames, with a shift of 64 samples between frames. The output signal was then truncated appropriately to match the length of the input signal, and normalized to match the power per unit time of the input signal.

Hou

The method of smearing created by Hou and Pavlovic (1994) attempted to simulate the reduced temporal resolution found in hearing-impaired listeners. As such, the frequency content of the processed signals was close to normal, and indeed, one of the intentions of the authors was to not corrupt the signal’s spectrum, as they intended to isolate the reduced frequency selectivity aspects of hearing loss while leaving out the other aspects of hearing loss. Also unlike the previous two models, this algorithm does not use an overlap-add paradigm, but instead first filters the signal into a number of adjacent frequency bands, then does processing on the entire waveform for each band. This was done to simulate the ear’s analysis of the sound by the “auditory filters” into similar adjacent frequency bands (or channels). By doing temporal smearing within each frequency region separately, the brain would be presented with a signal that would be similar to what hearing-impaired listeners’ brains would receive.

The model’s method of reducing temporal resolution was based on a model of human perception that included a few stages beyond the separation of the signal into different frequency channels. A block diagram of this model is shown in figure 2-11. In this model, after the signal is filtered into adjacent frequency bands, a nonlinear device transforms the signal. A squaring operation was assumed by the authors, although other nonlinear devices such as rectification are possible (Hou and Pavlovic, 1994). After the nonlinear stage is a temporal window that uses the information under
the window; and finally a decision device compares the resulting signals. The Hou and Pavlovic model incorporates stages to exploit the properties of this perceptual model, as is explained more in the following paragraph.

The specifics of the implementation of the Hou and Pavlovic (1994) processing method are as follows (see figure 2-12). The 10000 Hz vowel signals were first resampled to 16000 Hz, to have a sampling rate similar to the other smearing methods. Next, the signal is filtered into 32 channels by fourth-order Gammatone filters spaced one ERB apart. The Gammatone filter has a shape very similar to that of the roex(p) filter used in the Moore method. (The authors considered using roex(p) filters, which are presumably a better representation of the true auditory filters, but decided on Gammatone filters for their calculation efficiency.) For the purposes of implementing the described model as closely as possible, the current research also used Gammatone filters to create the channels for this stage. The impulse response of the Gammatone filter is

\[ I(t) = t^3 \exp(-2\pi bt) \cos(2\pi f_0 t), \]
Figure 2-12: Block diagram of the Hou temporal smearing model. The top diagram (a) shows the overall signal processing path, and the bottom diagram (b) shows the specifics of how temporal smearing is accomplished. Figure reproduced from Hou and Pavlovic, 1994.
"where $t$ is time, $f_0$ is the center frequency of the filter, and $b = 1.019B$ is the parameter determining the sharpness of the filter," with $B$ being the ERB of the filter (Hou and Pavlovic, 1994). In implementing the Gammatone filter bank, the paper used 23 filters covering a frequency span of 200-5000 Hz, as the stimuli they used were bandlimited to this frequency range. However, in the present implementation stimuli were used that had frequencies below 200 Hz, and (due to upsampling to 16000 Hz) up to 8000 Hz. Though the frequencies above 5000 Hz were generated through the upsampling process and extremely small in magnitude, to make the results as compatible as possible with the other smearing methods, filters were used up to this frequency.

After the signal is separated into channels, each channel is processed separately for smearing. The temporal smearing process, shown in figure 2-12 (b), first involved finding the temporal envelope of the channel, using a Hilbert transform. The envelope was then squared, and convolved with a "smearing temporal window." This smearing window has a shape similar to that of the auditory temporal window used in the perception model described above; both of these windows are roex($T_p$) functions. These have the form of

$$W(t) = (1 + 2t/T_p)exp(-2t/T_p),$$

where $t$ is time, $T_p$ is a time constant related to the window, and $ERD = 2T_p$, where $ERD$ is the equivalent rectangular duration of the window. In the paper, the value of $ERD$ ranged from 0-28ms. After the envelope of the signal was convolved with this temporal window, a square root was taken. Through this process of squaring the envelope, convolving it with the auditory temporal window, and taking the square root, the net effect on the signal was to generate a new auditory window that is the convolution of the auditory window internal to a listener and the smearing temporal window, if a normal-hearing listener heard the smeared signal:

$$W_{perceived\_new}(t) = W_{smearing}(t) * W_{internal}(t).$$
Following the square root, the original channel waveform is then scaled to have the new envelope (appropriately time-aligned). This preserves the fine structure but changes the envelope of the channel. Then, each channel is re-filtered with the same Gammatone filter that was used to generate the channel in the first place, in an attempt to reduce the spectral consequences of the time-domain smearing process. Finally, the smeared channels are added together again to produce the final output signal, and the result is normalized to the power per unit time of the original input signal. A sample of outputs from the model used by Hou and Pavlovic is shown in figure 2-13, while examples of outputs from the model used in the present simulation (which differs slightly from the original) are shown in figures 2-14 and 2-15.

2.5.2 Method of Recruitment

The method of loudness recruitment was taken from the paper by Moore and Glasberg (1993). The algorithm simulates the effects of threshold elevation found in hearing-impaired people, and the rapid increase in loudness level that occurs at audible levels. Loudness recruitment appears to stem from impaired-hearing listeners not having as much of the compressive nonlinearity normally found in the cochlea. Since a normal-hearing listener's cochlea will act to compress incoming sounds over a large range of levels, to induce the perception in normal-hearing listeners of not having this compression, the processing procedure expanded the envelopes of the signals exponentially. Also, in listeners with only one impaired ear, the loudness perceived in the impaired ear equals the loudness in the normal ear at a certain (high) level. This effect was also included in the simulation by not compressing portions of the signal above a threshold.

A general block diagram of the recruitment processing procedure can be found in figure 2-16. As can be seen in this diagram, the signal to be processed was first broken into a number of frequency bands. As with the similar step in the Hou smearing method, this step confined the effects of the processing so that the different "channels" would be processed independently. In this recruitment algorithm, the
Figure 2-13: Sample of outputs from the Hou temporal smearing model. The top panels (a and b) show the unsmeared signal and its spectrum, and the lower panels show the signal and its spectrum after it has been smeared with an ERD of 0.1ms and 28ms, respectively. Figure taken from Hou and Pavlovic, 1994.
Figure 2-14: Spectra resulting from different amounts of Hou smearing on the vowel II taken from the word BIIDH. All files were put through the recruitment algorithm but with no recruitment. The unsmeared signal is shown at the regular level, while the other traces have been offset by -50 dB increments. The labels to the left of the traces indicate the amount of smearing, that is, the equivalent rectangular duration of the smearing temporal window, in msec; US indicates the unsmeared signal.
Figure 2-15: Channels resulting from different amounts of Hou smearing on the vowel II taken from the word BIIDH. All files were put through the recruitment algorithm but with no recruitment. The unsmeared signal is shown at the regular level, while the other traces have been offset by -8 dB increments. The labels to the left of the traces indicate the amount of smearing, that is, the equivalent rectangular duration of the smearing temporal window, in msec; US indicates the unsmeared signal.
Figure 2-16: Block diagram of the model of loudness recruitment by Moore and Glasberg, 1993. The parallel arrows between the middle blocks represent the outputs of the 23 bandpass filters in the first block. Figure taken from Moore and Glasberg, 1993.

The filters used to create these channels had bandwidths similar to those of the auditory filters that would be found in impaired-hearing listeners. This was not to simulate additional frequency smearing, but to try duplicate the frequency independence that would be found in impaired-hearing listeners. Furthermore, this bandpass filtering allowed the amount of recruitment being simulated at different frequencies to be varied. Next, the envelope of each channel was found, and the fine structure was set aside for later use. The envelope of the signal was raised to a power $N$, then the fine structure was recombined with the now-recruited envelope. The purpose of expanding the envelope while leaving the fine structure left alone was to minimize spectral distortion. Finally, the channels were recombined to form a single output waveform.

The details of the processing are as follows. In the paper, the filters in the initial filter bank were each formed of a series of four first-order Gammatone filters. In the present implementation, however, it was decided to use a single fourth-order Gammatone filter as had been done in the Hou method. The simulation described in the paper used thirteen filters, with center frequencies from 100 to 5837 Hz. However, it was found that using the filters as described caused undesired spectral distortion in the resulting signal. So, instead, 23 filters were used, with center frequencies from 100 to 9190 Hz. The first six filters were spaced every 100 Hz, and thereafter they had a spacing of one ERB, with the ERBs corresponding to the auditory filters of a normal-hearing listener, $0.16f_{\text{center}}$.\(^1\) The filter center frequencies were the same as those

\(^1\)The 100-Hz spacing of the low-frequency auditory filters with non-constant ERBs probably contributed to some spectral distortion, but having logarithmically-spaced auditory filters there
used later in the $d'$ discrimination model, and the filter spacings were described in that paper (Farrar et al., 1987). Though the filters in the recruitment algorithm were spaced every ERB that a normal-hearing listener would have, the filters themselves had ERBs three times that width, corresponding to the ERBs of impaired-hearing listeners, as explained above. This led to a gain in the signal which was removed at the end of the processing for this algorithm.

After filtering into channels, the envelope of each channel was found via a Hilbert transform, and the fine structure was calculated by dividing the original signal by the envelope. Next, the envelope was filtered with a low-pass filter with a 3-dB point of 100 Hz. The purpose of this step was to prevent spectral distortion; also, Moore and Glasberg (1993) found that if the envelope were not low-pass filtered, the resulting sounds “had a somewhat ‘crunchy’ quality.” After the channel envelopes were low-pass filtered, each envelope was expanded in the following manner. It was assumed that the loudness recruitment would not take place above 100 dB, that is, above this level the loudness hearing-impaired listeners would experience would match the loudness experienced by normal-hearing listeners. If the unprocessed channel envelope, in dB, is $L_U$; the envelope after amplitude expansion occurs is $L_P$; and the power that the envelope is raised to (only below 100 dB) is $N$, then the following formula was used to simulate amplitude expansion:

\[
L_P = NL_U + K.
\]

$K$ is a constant defined by $K = (1 - N)100$ such that at 100 dB, $L_U = L_P = 100$ dB. Substituting in for $K$, we have:

\[
L_P = NL_U - (N - 1)100.
\]

If $L_U$ was above 100 dB, $L_P$ was set to equal $L_U$. This method of expansion was the one used in Nejime and Moore (1997), while the method in Moore and Glasberg would have been even more unnatural. The best solution would probably be to have fixed-width auditory filters with ERBs corresponding to an $f_{\text{center}}$ of 600 Hz.
multiplied the original Hilbert envelope by the low-pass filtered envelope raised to the power $N - 1$.

The current implementation assumed that a value of $N = 2$ corresponded to a 50 dB hearing loss. At this point, an attempt was made to fit the simulation's parameters to the audiograms of the different real subjects that took part in the confusion matrix experiment. To do this, for each subject, different values of $N$ were assigned to each channel in accordance with how much hearing loss the subject had at that frequency. Since loudness recruitment is caused by outer hair cell loss, the amount of which is correlated with the amount of hearing loss, the value of $N$ for the channel with center frequency $f_c$ was determined by the equation

$$N(f_c) = \frac{100}{100 - HL(f_c)},$$

where $HL(f_c)$ is the amount of hearing loss at frequency $f_c$, in dB.

Finally, the independently-recruited channels were added back together, and the output was scaled so that if the recruitment parameter was $N = 1$, i.e. no recruitment took place, then the level of the resulting spectrum matched the original signal's spectrum as much as was possible. The algorithm did modify the signal slightly even if no recruitment took place through the process of dividing the signal into channels and adding them back together again. Also, in the implementation by Moore and Glasberg (1993), the stimuli they used were low-pass filtered at 7000 Hz. Examples of the output from the current simulation are shown in figures 2-17, 2-18, and 2-19. As can be seen in these figures, the recruitment process produces both spectral distortion and absolute level shifts for signals below 100 dB, while for signals at 100 dB there is a little spectral distortion, and for signals well above 100 dB there is no distortion or change in level.
Channels corresponding to different amounts of recruitment on an originally 100 dB signal.

Figure 2-17: Channels resulting from different amounts of recruitment on the vowel II taken from the word BIIDH. The vowel was at a level of 100 dB. Recruitment of $N = 1.0$ (where $N$ is the exponent to which the envelope is taken) corresponds to a hearing loss of 0 dB, recruitment of $N = 2.0$ corresponds to a flat hearing loss of 50 dB, and in general the recruitment amounts correspond to hearing losses via the equation $N = 100 dB/(100 dB – hearinglossdB)$. No additional offsets were used in creating the plot: the effective level changes with increasing recruitment due to the increased absolute threshold.
Channels corresponding to different amounts of recruitment on an originally 80 dB signal.

Figure 2-18: Channels resulting from different amounts of recruitment on the vowel /i/ taken from the word BIIDH. The vowel was at a level of 80 dB. Recruitment of \( N = 1.0 \) (where \( N \) is the exponent to which the envelope is taken) corresponds to a hearing loss of 0 dB, recruitment of \( N = 2.0 \) corresponds to a flat hearing loss of 50 dB, and the other recruitment amounts correspond to hearing losses via the equation \( N = \frac{100dB}{(100dB - \text{hearing loss} dB)} \). No additional offsets were used in creating the plot: the effective level changes with increasing recruitment due to the increased absolute threshold.
Figure 2-19: Channels resulting from different amounts of recruitment on the vowel II taken from the word BIIDH. The solid lines are the result of recruitment with a simulated loss of 0 dB, and the dotted lines are the result of recruitment with a flat loss of 50 dB. The dashed line shows channels made from the original signal without recruitment, showing that simulating a loss of 0 dB does not distort the signal very much. Signals with levels of 120, 100, and 80 dB were recruited corresponding to losses of 0 dB and 50 dB, creating pairs of lines for each signal level. The 120 dB signal is above 100 dB the limit of recruitment, so recruitment has little effect. The 100 dB signal has the highest parts at around the limit of recruitment, so they are not affected, while the higher frequency portions of the signal are lower in level and so are distorted. The 80 dB signal is severely distorted and reduced in level by recruitment.
2.5.3 Combining the Smearing Methods and Recruitment

In the simulation, in some processing conditions the various smearing methods were applied in isolation, and in other processing conditions the model of recruitment was applied after one of the smearing methods had been applied. The case of smearing followed by recruitment was previously studied by Nejime and Moore (1997) using the same Moore frequency smearing function and model of recruitment used in the current simulation. Since the model of recruitment tended to modify the spectrum a little even if no recruitment was being done, in the current simulation all of the processing conditions were put through the recruitment model, but for the conditions where recruitment was not desired, a recruitment parameter of \( N = 1 \) was used. The eight different processing conditions done in this simulation are shown in table 2.5.3. In the table, an entry specifying recruitment indicates a value of \( N \) other than 1.

<table>
<thead>
<tr>
<th>First Stage</th>
<th>Second Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Moore smearing</td>
<td>Recruitment</td>
</tr>
<tr>
<td>ter Keurs smearing</td>
<td>Recruitment</td>
</tr>
<tr>
<td>Hou smearing</td>
<td>Recruitment</td>
</tr>
<tr>
<td>Moore smearing</td>
<td>—</td>
</tr>
<tr>
<td>ter Keurs smearing</td>
<td>—</td>
</tr>
<tr>
<td>Hou smearing</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 2.2: Table of the different processing conditions used in the simulation.

2.6 \( d' \) Discrimination Model

The next steps in the signal processing chain are part of the perceptual model. This model is based around the \( d' \) discrimination model (Farrar et al., 1987). In essence, the \( d' \) discrimination model assumes that people will discriminate between sounds based on their spectral differences. So, the model first finds a representation of the spectra of the signals to be compared. It assumes that signals will be classified based
on the linear differences between their spectra, then it computes a linear difference between pairs of spectra. Support for the model's accuracy is found in the Farrar et al. (1987) paper, in which listeners' responses to speech sounds modeled by shaped noise were reasonably well predicted by the model.

The model involved two sub-blocks, which are illustrated in figure 2-20. In the first block, the make channels block, the model filters the signal into a number of bands spaced by one ERB, as was done in the Hou and recruitment models. The center frequencies of the filters were the same as those used in the recruitment model. As with the Hou and recruitment models, this stage is meant to simulate the auditory filtering performed in humans. One difference from the Hou and recruitment models was the specific shape of the filters used to form the channels. The Farrar et al. paper used an earlier model of the filters, also derived by Patterson (1974). The form of these filters was

\[
H_j(f) = \left[ 1 + \left( \frac{f - f_{c,j}}{\alpha} \right)^2 \right]^{-1},
\]

where \( f_{c,j} \) is the \( j^{th} \) filter's center frequency, and \( \alpha \) is a constant that determines the width of the filter via \( ERB = (\pi/2)\alpha \). The \( ERB \)s of the filters were set to be 100 for the lowest 6 filters (up to 600 Hz), and were set to \( 0.16f_{c,j} \) for the filters centered at frequencies higher than 600 Hz. There were 23 filters total. After filtering, the total power per unit length was found for each filter output, then the natural log (ln) was taken of the result. The resulting quantity shall be referred to as the "channels" from a signal. It is important to note that though the signals used had different lengths, and even the ter Keurs processing method required a different sampling rate than the other methods, the process of finding the power per unit time accommodates for these differences. The steps just described filtering the signal into these channels took place in the make channels block in the main block diagram in figure 2-1, as well as in figure 2-20.

The next block in the \( d' \) discrimination algorithm compares the channels from
Figure 2-20: Block diagram of $d'$ discrimination model. The bottom block comes after the top block in the processing path. The top block is used to create channels from all vowels processed, and the bottom block takes all combinations of pairs of vowels, and produces a $d'$ distance between each pair. In the top figure, the signal is first filtered into bands. Then, in each band, the average power per unit length of the signal is found by first squaring it, summing over the signal length, dividing by the signal length, and then normalizing to account for variation in signal length between stimuli. Finally, the natural log of the resulting number is taken to produce a channel. In the bottom diagram, the difference ($\Delta_i$) between corresponding channels is found for a pair of stimuli, as indicated in the plot. These differences are then put through the formula in the bottom of the figure. $\Sigma$ is an arbitrary constant that is a scale factor for all $d'$ distances.
pairs of signals to find the $d'$ distance between each pair of signals. This is illustrated in the bottom part of figure 2-20. For a given pair of signals, the difference between each corresponding channel from the two signals was found, labeled $\Delta_i$ in the figure. Then, the $d'$ value was found via

$$d' = \sqrt{\sum_{i=1}^{23} \frac{\Delta_i^2}{\sigma^2}}$$

(2.12)

in which $\Delta_i$ and $\sigma$ have been previously described. At the output of this processing block, then, was a matrix of $d'$ values for all combinations of signal pairs.

One additional processing step was done while finding $d'$ values for pairs of signals. To simulate the effect of the real subjects’ hearing thresholds, the audibility of the signals was computed, for each subject separately. The levels of the channels were first converted into physical sound pressure levels through a gain. Next, this physical level was compared against the real subjects’ absolute hearing thresholds. In the $d'$ discrimination algorithm, if the level of any channel from either signal being compared fell below a subject’s hearing threshold, the difference between that channel and its pair was removed from the calculation of $d'$ for that subject. Or in other words, a simulation was done for each real subject, and the audibility of the stimuli to that subject was computed. In doing this, it was hoped that the results found for each subject would match the subject’s actual confusion matrix.

Finally, it should be noted that while the $d'$ discrimination model assumed that humans discriminate between sounds on the basis of the spectrum alone, the Moore processing method (and to some extent the ter Keurs processing method) only attempted to simulate the spectral effects of reduced frequency resolution in the impaired ear. So, while this perceptual model may be leaving some elements of the signals out relative to how real people would perceive sounds, it is entirely appropriate for evaluation of these frequency smearing models. The Hou model, however, attempted to simulate reduced timing resolution, and in fact tried to limit the spectral consequences. There were time differences between the lengths of the
vowels in the CVC tokens, so the real subjects’ comprehension would be affected by temporal smearing, but in the present simulation this was not taken into account.

2.7 Decision Model

Once the $d'$ values for all of the pairs of signals had been found, some way was needed to convert these results into a confusion matrix, for comparison with the real subjects' results. A decision model was used that assumed the following. Each stimulus generates a vector of perceptual cues in a perceptual space internal to a listener. The locations of these points in the perceptual space will be referred to as “stimulus centers.” The listener also creates a set of “response centers,” points in the perceptual space that are used as models or prototypes for each category of stimulus. In the present simulation, it is assumed there is one response center for each vowel, while the stimulus centers are the many examples of each vowel that are presented to a listener. It is assumed that the response centers are used in classifying the stimulus centers (or stimuli) into different categories. It is also assumed that, due to uncertainty in a listener’s mental representation, the cue vector for each stimulus center is the center of a multi-dimensional Gaussian probability distribution. The distribution corresponds to the various cue vectors that could be used by a listener in the actual decision of classifying that stimulus, and the distribution probabilities being the associated probabilities of those cue vectors being used. The last step of the decision model assumes that a stimulus will be classified in the category corresponding to the response center that its cue vector is closest to. A simple illustration of the model is shown in figure 2-21.

A block diagram of the inner workings of the decision model is shown in figure 2-22. In the implementation of this model, first the matrix of $d'$ values was put into a classical multi-dimensional scaling algorithm that produced a map of the points that accurately represented their $d'$ distances. Only the first three dimensions of the resulting map were used, but these provided sufficient accuracy. It was assumed
Figure 2-21: Illustration of decision model function in two dimensions. Stimulus centers are represented by the small points in the figure, and are surrounded by Gaussian distributions in perceptual space. Response centers are found by taking the centroid of the stimulus centers of a single type. The probability distributions around each stimulus center are assumed to evoke responses corresponding to the closest response center. The lines dividing the figure are the boundaries of these so-called response regions.
that this resulting map was the perceptual representation of the stimuli by a listener: the vector of coordinates of each point in this multi-dimensional map was the cue vector for that point in a perceptual space, and the dimensions of the map were the dimensions of the perceptual space.

To implement the perceptual uncertainty of each stimulus center, each stimulus center was replaced by a multi-dimensional Gaussian centered at the stimulus center's location. Also, the locations of the response centers were assumed to be the centroids of the stimulus centers for each type of vowel. Finally, each point in the Gaussian distributions surrounding the stimulus centers was assigned to category based on which response center it was closest to. A confusion matrix was formed by summing the probabilities of the points for each stimulus category that were classified as each of the possible responses.

In the simulation, the following parameters were used in this model: only the first 3 dimensions of the perceptual ($d'$) map were used, so the Gaussians were in 3 dimensions. A variance of $d' = 0.1$ was assumed, where the maximum dimensions of the $d'$ map were usually about 4-5 $d'$ units in the first dimension and about 3-4 $d'$ units in the second dimension, to give an idea of the scale.
Chapter 3

Results

Tables of simulated confusion matrices for all processing conditions are in Appendix B, and tables of the real subjects’ actual confusion matrices are found in Appendix C.

3.1 Total Error Rates

Plots of the total error rates for the simulated confusion matrices fitted to the different subjects’ data can be found in figures 3-1 through 3-10. Figures 3-1 through 3-5 show the total error rates of the simulated confusion matrices, while figures 3-6 through 3-10 show the error rates relative to the confusion matrices for unsmeared stimuli. All figures in this chapter are located at the end in a dedicated section.

The total error rates of simulated confusion matrices are almost always higher than the total error rates of the confusion matrices produced by the real subjects in 17 out of 120 smearing conditions. Of these, 15 of them were either the unsmeared stimuli or processed with the Hou smearing condition, which tended to give the similar results as the unsmeared stimuli. Moreover, the unsmeared stimuli have higher error rates than the real subjects in two cases, in one case they are almost identical, and in the other two cases they are within 7 percent. This indicates that the total error rates are limited by the perceptual model, since that was the processing stage responsible for the unsmeared stimuli producing large total error rates.
If we subtract the errors in the unsmeared stimuli from the rest of them (that is, find the error rates relative to the unsmeared stimuli), then the error rates are comparable to those of the actual subjects in a significant portion of the conditions. However, subtracting the confusion matrices from the other processing conditions does not produce confusions identical to those made by real subjects, as will be discussed later.

With increasing smearing for the Moore and ter Keurs methods, the total error rate generally increases approximately linearly, but with a varying slope. As the smearing is increased by a factor of 2, there is a change in total error rate of between 0 and ~18 percent, with the majority of cases having a change of 5-10 percent. This change is independent of whether or not recruitment occurs. With the Hou smearing and no smearing, there is little or no increase in error rate as amount of smearing increases.

The other trend in the total error rates is that the error rates with smearing followed by recruitment were above the error rates for the same smeared conditions but with no recruitment in all except 1 of 60 cases. Recruitment added a total error rate of between 0 to 20 percent, with a typical change of about 10 percent. This also meant that the error rates with recruitment were typically far above the subjects' actual error rates.

### 3.2 Locations of Errors in Confusion Matrices

The particular errors in the simulated confusion matrices occur in locations generally different from the errors in the real people's confusion matrices. There is some overlap in the real and simulated confusions, but there are also many confusions that are different between the real subjects and the simulation. With the real subjects, errors are typically in a few locations, while in the simulations they are much more widespread. Sample confusion matrices from the processed conditions are shown in table 3.1. These trends can be seen in a visual inspection of the sample confusion
matrices. Also, the locations of the maximum errors in the simulated confusion matrices tended to be in the same positions regardless of the type and degree of smearing, or which subject’s parameters were used in doing the processing. Generally, as the amount of smearing increased or recruitment was added, the confusions in the existing locations increased in magnitude, while new confusions in other locations were added that generally had smaller magnitudes. Errors in this same set of locations also occurred in the confusion matrices for the unsmeared stimuli, showing that this pattern was likely a consequence of the perceptual model.

Plots showing how well the simulated confusion matrices correspond to the real confusion matrices can be seen in figures C-1 to C-15 in Appendix C, and a plot showing how well the average data in table 3.1 fits to the real confusion matrices can be seen in figure 3-11. These plots show a $\chi^2$ measure and a Mean Squared Error (MSE) measure of the goodness of fit of the simulated matrices to the real subjects’ matrices. The MSE measure was computed by

$$\text{(3.1)} \quad N \sum_{i=1}^{36} (O - E)^2,$$

where $N$ is the total number of stimulus presentations, $O$ is an entry in a real subject’s confusion matrix, $E$ is an entry in a simulated confusion matrix, and the sum is over all 36 entries in a confusion matrix. The $\chi^2$ measure was computed by

$$\text{(3.2)} \quad N \sum_{i=1}^{36} \frac{(O - E)^2}{E} \text{ for } E > 0.5,$$

where $N$, $O$, and $E$ are the same as in equation 3.1.

In the $\chi^2$ measure, any entries in the simulated confusion matrices less than 0.5 were excluded because very small non-zero values would cause excessive contributions to the measure, and zero values would cause values equal to infinity. Even so, if an entry in the simulated confusion matrix is small and the entry in the real confusion matrix is large, that element will contribute to the total $\chi^2$ measure disproportionally more than it probably should for a “fair” measure. For this reason, the MSE measure
is possibly better to use in evaluating the goodness of fits between the real and simulated confusion matrices, although this measure has the problem that if there are small entries in both the real and simulated confusion matrices, the contribution of those elements is fairly trivial. It is this reason that causes some values to be excessively small (as can be seen in the plots). The plots show the natural log of both of these measures.

In general, as can be seen in the $\chi^2$ plots in figures C-1 to C-15, fitting the recruitment and audibility parameters to correspond to one of the subjects' audiograms did not produce results that fit that subject better. Also, the $\chi^2$ values produced were generally quite large, indicating that the simulated confusion matrices did not fit the real data very well. Values ranged from about 20-50 for subjects JO and RC, and values of about 150 for the other subjects for the $\chi^2$ measure, and about 7-20 for the MSE measure (on the plot, the natural log of these values is shown). With the $\chi^2$ measure, the processing condition that produced the best results varied quite a bit, with almost all processing conditions producing the best fit for some subject’s real confusion matrix. Also, there does not seem to be any benefit of using recruitment or not using recruitment to produce a better fit. With the MSE measure, in nearly all of the cases the unsmeared and Hou smearing without recruitment conditions produced the best-fitting confusion matrices. There seemed to be slightly better performance from the unrecruited conditions relative to the recruited conditions with this measure. Also, typically with both measures of goodness of fit the Moore method without recruitment did better than the ter Keurs method without recruitment for the same smearing multiplier.

Two of the subjects' (JO and RC) confusion matrices tend to fit the predictions better than did the other subjects' confusion matrices, regardless of the subject’s audiogram the recruitment and audibility parameters in the simulation were fit to. This was true especially for the unsmeared and Hou smearing with no recruitment conditions. This effect occurred because the real confusion matrices for these subjects had small total error rates, and the errors that did occur were spread out over many
low values rather than having a few large values, which is what happened with the other subjects’ real confusion matrices. Since the simulated confusion matrices tended to produce confusions in many locations, the contribution towards the $\chi^2$ measure from each of the confusions was smaller, but there were more such contributions. The end result was a smaller value of $\chi^2$.

Also, because the simulated confusion matrices corresponding to the unsmeared and Hou without recruitment conditions typically had the smallest error rates among the simulated confusion matrices, and because the simulated confusion matrices tended to have confusions in the wrong locations, these conditions tended to fit the real data better. The Hou method does not cause much change in the spectrum, as expected, so the confusion matrices produced by it are not much different than the unsmeared confusion matrices. Also, and notably, the unsmeared and Hou without recruitment conditions usually had total error rates closest to the total error rates of the real confusion matrices as well. These conditions fit subjects JO’s and RC’s confusion matrices particularly well due to the combination of both confusion matrices having small error rates, which tended to contribute less to the $\chi^2$ measures as described above.

If the confusion matrices resulting from the unsmeared conditions were subtracted from the other simulated confusion matrices, there was still no benefit in accuracy, as the resulting confusions still tended to be in the same locations.
<table>
<thead>
<tr>
<th></th>
<th>No Recruitment</th>
<th></th>
<th>With Recruitment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AH</td>
<td>EE</td>
<td>EH</td>
<td>II</td>
</tr>
<tr>
<td>No Sm.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AH</td>
<td>99</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>EE</td>
<td>0</td>
<td>90</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EH</td>
<td>2</td>
<td>0</td>
<td>92</td>
<td>2</td>
</tr>
<tr>
<td>II</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>84</td>
</tr>
<tr>
<td>OO</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>UH</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Moore Sm.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AH</td>
<td>94</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>EE</td>
<td>0</td>
<td>85</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>EH</td>
<td>7</td>
<td>0</td>
<td>76</td>
<td>5</td>
</tr>
<tr>
<td>II</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>60</td>
</tr>
<tr>
<td>OO</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>UH</td>
<td>2</td>
<td>0</td>
<td>11</td>
<td>19</td>
</tr>
<tr>
<td>ter Keurs Sm.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AH</td>
<td>88</td>
<td>0</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>EE</td>
<td>0</td>
<td>82</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>EH</td>
<td>9</td>
<td>0</td>
<td>62</td>
<td>10</td>
</tr>
<tr>
<td>II</td>
<td>0</td>
<td>4</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>OO</td>
<td>0</td>
<td>16</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>UH</td>
<td>3</td>
<td>0</td>
<td>21</td>
<td>16</td>
</tr>
<tr>
<td>Hou Sm.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AH</td>
<td>99</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>EE</td>
<td>0</td>
<td>90</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>EH</td>
<td>3</td>
<td>0</td>
<td>87</td>
<td>5</td>
</tr>
<tr>
<td>II</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>78</td>
</tr>
<tr>
<td>OO</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>UH</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 3.1: Table of the average simulated confusion matrices for each processing condition. Averages were taken across subjects and smearing amounts. From top to bottom, the processing conditions are: No smearing, Moore smearing, ter Keurs smearing, and Hou smearing. The left column contains the results of smearing without recruitment, and the right column contains the results of smearing followed by recruitment.
3.2.1 Comparison of Moore and ter Keurs Methods

Calculations showed that widening the ERBs by 3 times in the Moore method was the same bandwidth as smearing by 0.75 octaves in the ter Keurs method. However, the ter Keurs method almost always produced higher error rates than the Moore method, by about 5-15%. This can be easily seen on the graphs of total error rate (figures 3-1 to 3-5) since the points corresponding to 3 times the width and 0.75 octaves of smearing were set to be the “base” conditions with the smearing multiplier equal to 1. As the smearing multiplier increases the amount of smearing remains about the same for the two methods.

A table showing the average confusion matrices for the Moore and ter Keurs smearing methods with approximately the same total error rates is shown in table 3.2. As can be seen in this table, there are a few differences in the amounts of smearing caused by each method on the average, though the locations at which confusions occur are extremely similar. In confusing EH and II with UH, on average the Moore method produces larger II-UH confusions than EH-UH confusions, while the ter Keurs method produces the reverse. Also, the ter Keurs method produces larger 11-00 confusions than the Moore method on average. These are the primary differences between these average confusions created by smearing, although other minor differences remain. Though these differences can be seen in the averages, in the individual subjects the results vary to a large degree, with in some cases the reverse being true.

If the Moore and ter Keurs smearing methods (without recruitment) are compared with approximately equal total error rates, it is found that in almost all of the cases, with both measures of the goodness of fit, the ter Keurs smearing method gives a worse fit to the real subjects’ data than the Moore method. This can be seen in figures 3-12 and 3-13, which show the fitting measures compared with different smearing amounts but the other parameters the same. In the figures the ter Keurs results (dotted line) are almost always above the Moore results (solid line).
Table 3.2: Table of the simulated confusion matrices for approximately equivalent total smearing amounts for the Moore and ter Keurs methods, without recruitment. The results shown in the table are the averages across all subjects. The top row contains a Moore smearing with a widening factor of 4.5, and ter Keurs smearing of 0.75 octaves. The total error rate with that amount of smearing for the Moore method was a little lower than the total error rate for that amount of ter Keurs smearing. The bottom row contains a Moore smearing with a widening factor of 6, and ter Keurs smearing of 1.125 octaves. The total error rate with that amount of smearing for the Moore method was a little higher than the total error rate for that amount of ter Keurs smearing.

### 3.3 Effects of Processing Parameters on the Results

This section will discuss how altering different parameters in the processing affected the resulting confusion matrices.

#### 3.3.1 Loudness recruitment and Audibility

The effect of the loudness recruitment algorithm is mainly to lower the level of the spectrum and linearly increase the level differences between different parts of the spectrum. Recruitment also tends to introduce some distortion in the signal, causing its spectrum to be smoothed out similarly to the effect of the frequency smearing.
However, this effect occurs to the largest extent in the high-frequency parts of the signal. These high frequency portions of the signal are frequently not audible in the current simulation, thereby removing this effect of recruitment. A plot showing the percent of stimuli that had a given channel audible vs. the channel number is shown in figure 3-14. Because of this, the main effects of recruitment in this simulation were likely to alter the audibility of the different channels, and to increase the $d'$ distances through expanding the spectrum.

Also, for two of the subjects (JO and MG), the stimuli were presented at a high level (>105 dB) such that recruitment did not make as much of a difference, only causing a change in total error rate of up to about 8% for the Moore and ter Keurs smearing methods, and about 10% for the Hou and unsmeared conditions. The smaller influence of recruitment can be also seen through inspecting the confusion matrices associated with these subjects, in Appendix B.

One additional note should be made about the audibility of the high frequencies. It is clear there is something wrong with the audibility in the current simulation. In the original experiment with real subjects, the levels of high- and low-passed sounds with a cutoff of 2500 Hz were adjusted to maximize the subjects’ comprehension of the syllables being played. The current simulation predicts that, at the ideal levels found in the previous experiment, the high-passed portion of the signal would have likely been entirely inaudible for all of the subjects. This was clearly not the case, so something is probably wrong with the simulation. To accommodate for this, a second simulation was done that assumed all frequencies were audible for all subjects. The results of this are discussed later.

### 3.3.2 Male vs. Female Speakers

In the simulation of the perceptual model, the stimuli corresponding to the male speaker and the female speaker formed two quite distinct clusters in the perceptual space (of $d'$ distances) map. As such, the male and female speakers’ stimuli were put through the perceptual model (and classified) separately. With the male and
female speaker stimuli being classified separately, to form a single confusion matrix at the end the confusion matrices for the male speaker and female speaker were averaged. This would accurately reflect the responses of a person exposed to the entire stimulus set, as approximately half of the vowel samples were spoken by the male speaker, and half were by the female speaker. The confusion matrices produced by the simulation with the male and female points separated were more accurate (as determined by the $\chi^2$ and MSE measures) than if the stimuli spoken by the male and female speakers were mixed together in the perceptual model. This was largely due to the fact that when they were classified together, the distribution of stimulus centers corresponding to each vowel tended to be much larger than the distribution of stimulus centers from either the male or female speaker separately. The larger distribution of stimulus centers resulting from the combined male and female simulation, as well as the distinct clusters formed in the perceptual map of the stimuli from the two speakers, are illustrated in figure 3-15. A table of the averages across subjects, with a smearing multiplier of 1, is shown in table 3.3. In this table, the amounts of confusions are larger than with the averages from the male and female speaker stimuli classified separately, despite the fact that the smearing multiplier of 1 data corresponded to the least error rates in the separate simulation. A plot showing the $\chi^2$ and MSE measures of how well the data from the table fit the real subjects’ confusion matrices is in figure 3-16. In the figure, the MSE measure and the $\chi^2$ measure for subjects JO and RC shows that this combined simulation has a much poorer fit than if the male and female speakers’ stimuli are classified separately and then combined. The other subjects for the $\chi^2$ measure show comparable results.

### 3.4 Results with Simulating Full Audibility

In this new simulation, it was assumed that all frequencies were audible to all subjects; so, all channels were included in calculating the $d'$ distance between pairs of vowels.
3.4.1 Error rates

Graphs of the total error rate assuming that all frequencies are audible are shown in figures 3-17 through 3-21. In these plots, the points corresponding to the un-recruited signals are the same in every plot, because the only difference between them was the audibility of the points, and now that difference has been removed.

The error rates of the unrecruited stimuli with full audibility decreased or stayed about the same as the error rates of the reduced-audibility stimuli. This shows that decreasing the audibility of high frequencies has a detrimental effect on perception.

Even with the reduced error rates, the subjects’ actual error rates still tended to be below the calculated error rates. The results of the ter Keurs smearing method were never below the actual error rates; with small amounts of smearing, however, the Moore method results could match or be close to the actual error rates. Also, in all subjects, the unsmeared confusion matrices were below or close to the actual error rate, showing that it could be theoretically possible to apply a smearing algorithm and match the error rate exactly. The error rates with recruitment were usually far above the subjects’ actual error rates, as with the reduced audibility case.

Also as with the reduced audibility simulation, the error rates for the Moore and ter Keurs smearing methods increased with the amount of smearing approximately linearly, while the error rates for the Hou and unsmeared conditions did not increase.

Comparing recruitment with and without all frequencies being audible, the error rates corresponding to the recruited signals generally stayed the same or decreased a little with full audibility relative to the simulation with reduced audibility. Although increasing the audibility will tend to decrease the error rate (as shown by the unrecruited stimuli), making use of all the frequencies includes the frequencies that were smeared by recruitment, which would tend to increase the error rate. So, it is hard to conclude anything based on this finding. Also, since the audibility differences have been removed, the only difference in the recruited signals between subjects is the amount of recruitment applied at different frequencies (according to the subjects’ audiograms). It can be seen in the figures that the between-subject error rates with
recruitment can vary by up to 20 percent, showing that recruitment is a significant factor in producing confusions when the affected frequencies are audible.

3.4.2 Locations of Errors

In general, with full audibility the locations of the confusions (and maximum confusions) were usually the same as those with reduced audibility. The amounts of the confusions produced did vary relative to the reduced-audibility case, however. Tables showing the average confusion matrices for the different processing conditions and comparing the Moore and ter Keurs smearing results for similar total error rates are in tables 3.4 and 3.5, respectively. A plot showing the $\chi^2$ and MSE measures for how well the average data in table 3.4 fits the real subjects’ confusion matrices can be seen in figure 3-22. As with the reduced audibility case, altering the parameters of the simulation to fit one of the subjects better than the others did not produce results that fit that subject better.

Also, the plots showing the $\chi^2$ and MSE goodness of fit measures were very similar to those corresponding to the reduced-audibility case. These can be seen in Appendix C in figures C-16 through C-30. The values in these figures are also large, similarly to the reduced audibility condition. With both goodness-of-fit measures, using recruitment tended to produce worse results than not using recruitment in the majority of cases, but this was not true in general. As with the reduced-audibility case, the unsmereared and Hou smearing (without recruitment), and subjects JO and RC tended to produce better fits than the other conditions. In general, the goodness of fits with the $\chi^2$ measure tended to be better with the all-frequencies-audible condition, while with the MSE measure the goodness of fits was about the same.

3.4.3 Comparison of Moore and ter Keurs Methods

With the full-audibility condition, if the Moore and ter Keurs smearing methods (without recruitment) are compared with approximately equal total error rates, it is
found that neither method consistently gives better results than the other. This is different from the reduced-audibility case, in which the Moore method was better. This can be seen in figures 3-23 and 3-24, which show the fitting measures compared with different smearing amounts but the other parameters the same.
Table 3.3: Table of the average simulated confusion matrices for each processing condition. Averages were taken across subjects for a smearing multiplier of 1. From top to bottom, the processing conditions are: No smearing, Moore smearing, ter Keurs smearing, and Hou smearing. The left column contains the results of smearing without recruitment, and the right column contains the results of smearing followed by recruitment.
Table 3.4: Table of the average simulated confusion matrices for each processing condition, assuming that all frequencies are audible. Averages were taken across subjects and smearing amounts. From top to bottom, the processing conditions are: No smearing, Moore smearing, ter Keurs smearing, and Hou smearing. The left column contains the results of smearing without recruitment, and the right column contains the results of smearing followed by recruitment.

<table>
<thead>
<tr>
<th></th>
<th>No Recruitment</th>
<th>With Recruitment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AH</td>
<td>EE</td>
</tr>
<tr>
<td><strong>No Sm.</strong></td>
<td>AH</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>EE</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>EH</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>OO</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>UH</td>
<td>1</td>
</tr>
<tr>
<td><strong>Moore Sm.</strong></td>
<td>AH</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>EE</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>EH</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>OO</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>UH</td>
<td>2</td>
</tr>
<tr>
<td><strong>ter Keurs Sm.</strong></td>
<td>AH</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>EE</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>EH</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>OO</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>UH</td>
<td>2</td>
</tr>
<tr>
<td><strong>Hou Sm.</strong></td>
<td>AH</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>EE</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>EH</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>OO</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>UH</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3.4: Table of the average simulated confusion matrices for each processing condition, assuming that all frequencies are audible. Averages were taken across subjects and smearing amounts. From top to bottom, the processing conditions are: No smearing, Moore smearing, ter Keurs smearing, and Hou smearing. The left column contains the results of smearing without recruitment, and the right column contains the results of smearing followed by recruitment.
Table 3.5: Table of the simulated confusion matrices for approximately equivalent total smearing amounts for the Moore and ter Keurs methods, without recruitment, assuming that all frequencies are audible. The results shown in the table are the averages across all subjects. The top row contains a Moore smearing with a widening factor of 4.5, and ter Keurs smearing of 0.75 octaves. The total error rate with that amount of smearing for the Moore method was a somewhat (~5%) lower than the total error rate for that amount of ter Keurs smearing. The bottom row contains a Moore smearing with a widening factor of 6, and ter Keurs smearing of 1.5 octaves. The total error rate with that amount of smearing for the Moore method was almost identical to the total error rate for that amount of ter Keurs smearing.
Figure 3-1: Total error rates for the simulation using subject JB’s audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject JB’s actual error rate.
Figure 3-2: Total error rates for the simulation using subject JO’s audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject JO’s actual error rate.
Total error rates for subject MG

Figure 3-3: Total error rates for the simulation using subject MG’s audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject MG’s actual error rate.
Figure 3-4: Total error rates for the simulation using subject PW’s audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject PW’s actual error rate.
Figure 3-5: Total error rates for the simulation using subject RC’s audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject RC’s actual error rate.
Figure 3-6: Error rates relative to the total error rate for unsmereared stimuli for the simulation using subject JB’s audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject JB’s actual error rate.
Relative error rates for subject JO

No smearing
Moore sm.
ter Keurs sm.
Hou sm.

Figure 3-7: Error rates relative to the total error rate for unsmeared stimuli for the simulation using subject JO's audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject JO's actual error rate.
Figure 3-8: Error rates relative to the total error rate for unsmeread stimuli for the simulation using subject MG's audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject MG's actual error rate.
Relative error rates for subject PW

Figure 3-9: Error rates relative to the total error rate for unsmeared stimuli for the simulation using subject PW’s audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject PW’s actual error rate.
Figure 3-10: Error rates relative to the total error rate for unsmeared stimuli for the simulation using subject RC’s audiogram. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject RC’s actual error rate.
Figure 3-11: Plot showing how well the average of all of the confusion matrix data fit to the subject's actual confusion matrices. Averages were taken across all subjects and smearing amounts, for the reduced-audibility simulation.
Comparison of Moore and ter Keurs smearing with about equal error rates: $\ln(\text{Chi-squared})$ measure

$\ln(\text{Chi-squared value})$

Different subjects fitted. Moore smearing of $S.M. = 1.5$, ter Keurs smearing of $S.M. = 1.0$

Comparison of Moore and ter Keurs smearing with about equal error rates: $\ln(\text{O-E})^2$ measure

$\ln(\text{O-E}^2\text{value})$

Different subjects fitted. Moore smearing of $S.M. = 1.5$, ter Keurs smearing of $S.M. = 1.0$

Figure 3-12: A comparison of the ter Keurs results (dotted line) and Moore results (solid line) if the total error rates are approximately the same for the two smearing conditions. The ter Keurs result had a smearing multiplier of 1.0, and the Moore smearing had a multiplier of 1.5. For each pair of ter Keurs and Moore values (plotted in the same vertical space), the subject used to create the simulated matrix and the real subject it was being compared to were the same.
Comparison of Moore and ter Keurs smearing with about equal error rates: \( \ln(\text{Chi-squared}) \) measure

Different subjects fitted. Moore smearing of S.M. = 2.0, ter Keurs smearing of S.M. = 1.5

Comparison of Moore and ter Keurs smearing with about equal error rates: \( \ln(\text{O-E}\text{^2}) \) measure

Different subjects fitted. Moore smearing of S.M. = 2.0, ter Keurs smearing of S.M. = 1.5

Figure 3-13: A comparison of the ter Keurs results (dotted line) and Moore results (solid line) if the total error rates are approximately the same for the two smearing conditions. The ter Keurs result had a smearing multiplier of 1.5, and the Moore smearing had a multiplier of 2.0. For each pair of ter Keurs and Moore values (plotted in the same vertical space), the subject used to create the simulated matrix and the real subject it was being compared to were the same.
Figure 3-14: Graph showing the percentage of stimuli that had audible channels for a given channel number, for each subject. Note that the channels correspond to a log scale with channel 1 centered at 100 Hz, channel 12 at 1575 Hz, and channel 19 at 4839 Hz (for reference). Also, it can be seen that in most plots two groups of curves can be seen. The set of curves to the left corresponds to no recruitment, and the set to the right is with recruitment.
Figure 3-15: Plot showing the stimulus centers and response centers for the stimuli from the male and female speakers combined (top plot), and just the female speaker (bottom plot). The stimulus centers are shown by the small black dots, and the response centers are shown by the open circles. In each plot, the response centers are the average of the stimulus centers for each vowel. Only half of the total number of stimulus centers were plotted for computational reasons. As can be seen around the “AH” in the top plot, the stimuli corresponding to the male and female speakers formed two distinct clusters for each vowel. The clusters are not distinct for the other vowels in this plot due to the perspective. Also, it can be seen that the stimulus centers for each vowel are less spread out in the bottom plot.
Figure 3-16: Plot showing how well the average of all of the confusion matrix data fit to the subject's actual confusion matrices. Averages were taken across all subjects, with the smearing multiplier fixed at 1.0, for the reduced-audibility simulation.
Figure 3-17: Total error rates for the simulation using subject JB’s audiogram, assuming that all frequencies are audible to the subject. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject JB’s actual error rate.
Total error rates for subject JO, assuming all frequencies are audible

Figure 3-18: Total error rates for the simulation using subject JO's audiogram, assuming that all frequencies are audible to the subject. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject JO's actual error rate.
Figure 3-19: Total error rates for the simulation using subject MG’s audiogram, assuming that all frequencies are audible to the subject. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject MG’s actual error rate.
Figure 3-20: Total error rates for the simulation using subject PW’s audiogram, assuming that all frequencies are audible to the subject. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject PW’s actual error rate.
Figure 3-21: Total error rates for the simulation using subject RC’s audiogram, assuming that all frequencies are audible to the subject. Solid lines indicate results with just smearing and no recruitment, as indicated in the legend, and dotted lines indicate those conditions followed by recruitment. The dashed line is subject RC’s actual error rate.
Figure 3-22: Plot showing how well the average of all of the confusion matrix data fit to the subject's actual confusion matrices. Averages were taken across all subjects and smearing amounts, for the simulation in which all frequencies were audible.
Figure 3-23: Assuming full audibility, a comparison of the ter Keurs results (dotted line) and Moore results (solid line) if the total error rates are approximately the same for the two smearing conditions. The ter Keurs result had a smearing multiplier of 1.0, and the Moore smearing had a multiplier of 1.5. For each pair of ter Keurs and Moore values (plotted in the same vertical space), the subject used to create the simulated matrix and the real subject it was being compared to were the same.
Figure 3-24: Assuming full audibility, a comparison of the ter Keurs results (dotted line) and Moore results (solid line) if the total error rates are approximately the same for the two smearing conditions. The ter Keurs and Moore results both had a smearing multiplier of 2.0. For each pair of ter Keurs and Moore values (plotted in the same vertical space), the subject used to create the simulated matrix and the real subject it was being compared to were the same.
Chapter 4

Discussion

4.1 Why the Simulation Did Not Work

The primary problem in this simulation seemed to be that the locations of the largest confusions produced at the output of the perceptual model tended to occur in the same locations, regardless of the amount of smearing or even if there was any smearing at all. A certain set of confusions was present in the results for the unsmeared stimuli. As the amount of smearing increased or recruitment was added, the degree of confusions at these locations increased, and at other locations new confusions arose, but these were generally of a smaller magnitude than the original confusions. With recruitment, there tended to be confusions in almost all entries in the resulting confusion matrix, and even with no recruitment but large amounts of smearing there were many entries in the confusion matrices.

Real normal-hearing human listeners do not make many vowel confusions in quiet, and they do not even if they are listening to stimuli in quiet that have been smeared with large amounts of smearing (Baer and Moore, 1993, and ter Keurs et al., 1992). Because humans do not make confusions where the perceptual model in the simulation does, this implies that humans use a different algorithm for doing perception than the model. Either humans must use different cues than the model, or they must use the cues (i.e. the spectrum) in a different manner.
Vowel duration is another cue that can be used to discriminate vowels. Certainly, the vowels presented to the subjects in the real experiment differed in duration, as shown by figure 4-1. However, the standard deviation of these lengths was large, and this was almost certainly not the only difference between the real human perceptual model and the simulated perceptual model.

Instead, it is more likely that humans use the spectral cues of the vowels in a different way than the simulated perceptual model. This is easily illustrated by the diagrams in figure 4-2. On the left side of the diagram, there are three sample spectra, with peaks centered at two possible frequencies representing the formants in a vowel. Spectra 1 and 2 have a single peak at the first frequency, but the width and height of the formant varies. Spectrum 3 has a smaller peak at the first frequency, but also a peak at the second frequency. Because of the differences in the “formants,” the $d'$ algorithm used in the simulation would produce that the spectra were equidistant from each other in a perceptual space, shown in the bottom portion of the diagram. However, because of the presence of the second peak, a human would judge that the third spectrum would be far from the other two. This is illustrated in the perceptual map corresponding to a real human, in the bottom right of the diagram. So, in this manner, the $d'$ algorithm would classify the spectra differently than a real human, which would result in different confusions.

Another example, on the right side of the figure, shows a spectrum that is flat except for a single bump. In spectrum 1, the bump is wide and short, while in spectrum 2 it is narrower but taller, and in spectrum 3 it is even more narrow and tall so as to be tonelike. The $d'$ algorithm in the simulation would judge these to be equally different from each other, but a human would easily be able to judge that the tonelike stimulus was very different from the other spectra. It should also be pointed out that the $d'$ algorithm used in the simulation was developed and tested on broadband noise stimuli, and vowels and tones are not noise.

So, it is proposed that the following situation occurred in this simulation, that caused the results to be other than desired: regardless of how real people do
discrimination, the $d'$ algorithm used in this simulation should reflect the fact that some of the vowels' spectra got closer together as a result of smearing, and thus would be confused in the perceptual model. Assuming that the smearing models more or less correctly duplicate the smearing process that goes on in hearing-impaired listeners, the confusions that real people make should be reflected in the $d'$ distances between those vowels getting smaller. Let us assume for the moment that this actually happened in the current simulation.

Even with the $d'$ algorithm reflecting that the vowels humans typically confuse were getting closer together through smearing, another effect was also taking place in the current simulation. In particular, the $d'$ discrimination algorithm used in the simulation will confuse some vowels that real humans could easily tell apart, for example by their spectral pattern. An example of this was illustrated above. This effect most likely occurred in the current simulation, causing certain confusions to exist that humans would not have difficulty with due to the different discrimination algorithm used by humans. This caused the confusions to exist in the unsmeared stimuli.

When smearing was applied, these already-confused stimuli became more similar to each other, along with the stimuli that humans would be confused by. So, although the confusions made by real people were represented in the perceptual map (or confusion matrices), their effect was masked by the other confusions increasing in magnitude. In some cases, this effect could have occurred to such a degree that the simulation would predict the absence of a confusion between some stimuli, while a human would be confused.

So, because of this background of large numbers of confusions, it was very difficult to detect the real confusions. Notably, the smearing algorithms did almost-linearly increase the total error rate of the confusion matrices, so it is clear they were doing something that contributed towards the signals being confused with one another. But, it is unknown whether or not the assumption that they were correctly simulating the frequency smearing is correct, or if they were creating inaccurate confusions.
One way of testing whether or not this interpretation of the results is correct is to examine the directions that the stimuli moved on the average relative to one another. A necessary constraint for this hypothesis being correct is that if there are confusions produced by the real subjects, then the vowels involved must move closer together as smearing is increased. If the vowels moved further apart (i.e. the $d'$ distance increased), then either this explanation must be incorrect, the smearing and recruitment algorithms are not functioning properly, or some other mechanism besides outer hair cell loss is responsible for the real subject's hearing loss. Examining the relative motion of individual pairs of vowels is likely the only way to determine if the smearing/recruitment results are accurate or not, but even then there is still the effect that the confusions different than those humans would make would change and obscure the desired confusions.

The possibility of using vowel duration as a cue is worth discussing additionally. It could be the case that the real subjects were able to distinguish between sounds on the basis of the vowels' durations, even if their spectra were very similar. If this were the case, we would expect to see fewer confusions in those locations with very different durations in the real confusion matrices. Assuming that the extracted vowel durations are indeed reflective of the actual vowel durations perceived by the real subjects, vowel duration was probably not an important cue in vowel discrimination. Many of the real subjects made AH-II and AH-UH confusions, but the duration differences of these vowels are the largest among the stimulus set. It was also the case that vowels with similar durations did tend to be confused by the real subjects, and in some cases (i.e. OO-UH) the confusion magnitudes were quite large, which would be consistent with vowel duration being used as a cue. However, the existence of confusions between vowels of very different durations argues against vowel duration being an important cue.
4.2 Other Sources of Error

Besides the perceptual model operating differently from how real humans do perception, there were several other possible sources of error that would cause the simulation results to not duplicate the real subjects' confusion matrices well.

First, the real subjects' confusion matrices differed somewhat in terms of the location and severity of confusions, though there were a number of similarities. Many of the locations at which confusions occurred could be found in as many as three or four out of the five subjects (see Appendix A for the actual confusion matrices). However, some variation still remained. It seems reasonable to hypothesize that the mechanisms of hearing loss in the different subjects varied to some extent because of this. Other factors that could have varied between subjects include the amount of outer hair cell versus inner hair cell loss and possible neural conduction losses in the auditory nerve. In the current simulation, it was assumed that all of the hearing loss was due to outer hair cell loss. To the extent that this assumption was not true, there could have been differences between the simulated and real confusion matrices. Also, the current simulation simulated a constant widening of the auditory filters for all subjects (which would correspond to a flat loss), regardless of the subjects' audiograms. While the assumption that the subjects' audiograms are flat is approximately true, it is definitely not the case for subjects MG and RC. It is interesting to note that despite this assumption not being true, subject RC's real confusion matrix correlated better with the simulated confusion matrices than the other subjects' (besides JO's) confusion matrices. This lends support to the earlier explanation that it was the lower total error rates and distribution of errors that caused these subjects' confusion matrices to fit the simulated matrices better.

4.3 Conclusions that Can Be Drawn

Despite the fact that the simulated confusion matrices tended to not have their largest confusions in the same locations as the real confusion matrices, some conclusions
still can be drawn from the results. The smearing and recruitment algorithms still introduced errors into the confusion matrices, indicating that something was occurring. The confusion matrices are reflective of the amount of frequency smearing that occurred through the processing, because the perceptual model acted as a sort of spectrum analyzer.

4.3.1 Total Error Rates and Smearing and Recruitment

It is interesting to compare the exact error rates produced by this simulation with the error rates found when Baer and Moore (1993), ter Keurs et al. (1992), and Moore and Glasberg (1993) simulated smearing and recruitment, respectively, and presented the results to normal-hearing listeners.

Baer and Moore (1993) presented the results of frequency-smeared words to listeners at 65 dB SPL with a background of quiet. With No Smearing, ERB x3, and ERB x6 conditions, the total error rates were about 0% for the first two, and 1.7% for the third condition. This is very low compared to the total error rates found in the current simulation with Moore smearing and no recruitment, which were in the 10-30% range for smearing of ERB x3, and in the 25-40% range for the ERB x6 condition. Also, in the current simulation as the frequency smearing was increased from ERB x3 to ERB x6 the total error rate increased about 10-15% on average. This is far more than the change found by Baer and Moore. One factor that could potentially explain this difference is that in the Baer and Moore experiment, listeners heard entire words, while in the current experiment vowel identification in nonsense syllables was studied. Listeners in the Baer and Moore experiment could potentially combine cues from different phonemes in the entire word to improve comprehension.

ter Keurs et al. (1992) simulated frequency smearing with isolated vowels in quiet, that had been equalized in length, and were presented at 65 dBA. They only performed smearing at 1/8, 1/2, and 2 octave amounts, and found that with the unsmeared, 1/8, and 1/2 octave conditions, the total error rate was only 3%, but with 2 octave smearing it jumped to 68%. In comparison, in the current simulation
only the smearing amounts of 0.75, 1.125, and 1.5 octaves was investigated (without recruitment). For all subjects, the total error rates were about 25-35% at 0.75 octave smearing, about 30-37% for 1.125 octave smearing, and about 33-40% for 1.5 octaves of smearing. These error rates are clearly between 3% and 68%, but it should also be noted that in the simulation, the unsmeared vowels had a total error rate of 6-20%. Because the smearing amounts in the current simulation did not overlap well with the smearing amounts in the ter Keurs et al. (1992) experiment, it is difficult to form conclusions about the accuracy of the simulation based on just the total error rates. Also, in the ter Keurs et al. experiment, they used 11 different vowels, while the simulation and experiment by Brantley only used 6.

However, ter Keurs et al. (1992) also formed a confusion matrix based on the subjects’ responses to the 2-octave simulation. A comparison of this result with the current simulation is justified, as one would expect the confusions to be approximately in the same locations in both cases, even though the confusion amounts may be different. ter Keurs et al. found that almost all of the confusions were vowels being judged by subjects as being back vowels. The experiment used all of the vowels in the current simulation except UH, but it instead used “AW” which is similar. They also found a large confusion with EH being classified as AH, which also appeared in the current simulation, and they found minor confusions with II being classified as AH and EH, the former of which was not found in the current simulation, but the latter was. In the current simulation, there do exist many confusions with vowels being classified as back vowels, with the exception of EE-UH, AH-OO, and EH-OO. This suggests that the simulation may be somewhat reflective of the responses of human listeners. However, the simulation also produced large confusions at other locations, most notably II-EE, OO-EE, AH-EH, EE-EH, EH-II, OO-II, UH-EH, and UH-II. This large amount of other confusions limits the degree to which it can be concluded that the simulation and experiment produced the same results.

Moore and Glasberg (1993) processed sentences with the recruitment algorithm and presented them to subjects, using different input levels and different amounts of
recruitment. It is unfortunately hard to compare the results of their experiment with the current simulation, because they only presented the 50% correct points in the paper. With a severe flat hearing loss simulated, the 50% correct point was with an input level of about 79 dB SPL, and with a moderate-to-severe sloping loss simulated, the 50% correct point was with an input level of about 52 dB SPL. In comparison, in the current simulation subjects MG and RC had an approximately moderate-to-severe sloping loss. For subject MG, the input level was about 108 dB for both high and low frequencies, and simulating recruitment produced a total error rate of about 14%. For subject RC, the input level was 80 dB SPL up to 2500 Hz, then 102 dB SPL above 2500 Hz. This produced an error rate of nearly 30%. In both of these cases, the input levels were higher and the resulting error rates were lower than those in the Moore and Glasberg experiment. This is consistent with the simulation operating correctly, but it is hard to tell to what extent.

Also in the current experiment, subjects JB, JO, and PW had approximately flat moderate to severe losses. Subjects JB and PW had the input signals present at levels of (80,95) and (85,88) dB SPL, where the numbers in parenthesis indicate the levels below and above 2500 Hz, respectively. In response, subject JB’s total error rate was about 40%, and subject PW’s was about 27%. For these subjects, the error rate for unsmeared stimuli was 21% and 12%, respectively, indicating that this contributed somewhat to the recruited signals’ total error rates. Even so, the results for subject JB are roughly comparable to the results found in the Moore and Glasberg experiment. The results for subject PW are consistent with recruitment causing increased error rates, but cannot specify to what degree. For subject JO, the signal was presented at a high level (105,110) dB SPL, causing the error rate to be only 15%.

In general, it was found in the simulation that the total error rate as a result of recruitment was comparable to the total error rates for Moore smearing of ERB x3 or ERB x4.5, and usually less than the total error rates for all amounts of ter Keurs smearing. In comparison, the results of the actual experiments indicate that Moore smearing should cause a very low error rate and the error rate for recruitment
be greater than it. Moreover, in the current experiment, the input levels to the recruitment algorithm were larger than those used in the actual experiments, but the resulting error rates were still larger than those corresponding to the Moore smearing. This likely indicates that the effect of the perceptual model was to drastically increase the error rates due to Moore smearing, and perhaps also increase the error rates due to recruitment.

Comparing the results of the Moore and ter Keurs smearing, the Moore results in the simulation had lower error rates than those from the ter Keurs smearing for comparable amounts of smearing. This result from the simulation is most likely consistent with the experimental results from Baer and Moore, and ter Keurs et al., although in the ter Keurs et al. (1992) experiment, no data is available for smearing bandwidths between 1/2 and 2 octaves.

Comparing the results of the ter Keurs smearing and recruitment, recruitment almost always caused total error rates below those from the ter Keurs smearing. Given that in the simulation the input levels to the recruitment algorithm were larger than those used in the experiment by Moore and Glasberg, which would tend to reduce the error rates, the simulation results are likely consistent with the experimental results, although again data for the ter Keurs et al. experiment is not known for intermediate smearing amounts.

4.3.2 Comparing Moore and ter Keurs Smearing

In the reduced audibility condition, it was found that the results of the Moore smearing fit the real subjects’ confusion matrices better than did the results of the ter Keurs smearing, for similar total error rates. It is not obvious why this occurred. The Moore method usually produced slightly fewer confusions than the ter Keurs method, and had larger confusions in a few locations. This may have caused a few large elements in the real confusion matrices to be excluded from the $\chi^2$ calculation, although the MSE measure also indicated that the Moore method was better, although less strongly. In any event, because of this result it can tentatively
be concluded that (without recruitment) the Moore smearing method produces more accurate results than the ter Keurs smearing method. This result involved the comparison of all of the simulated matrices to all of the real subjects' matrices, so it does not necessarily indicate that the Moore method is better at creating the specific confusions made by a particular subject, if the parameters in the model are set to correspond to that of the subject, even though this was true; rather, it indicates that the Moore method is better at creating confusions that are generally made by hearing-impaired listeners. However, it should be noted that though the Moore method produced better fits than the ter Keurs method, the fits still were not very accurate.

4.3.3 The Effect of Audibility

In comparing the results between the reduced-audibility and full-audibility simulations, it is clear that the audibility of the different parts of the spectrum makes a great deal of difference for unrecruited stimuli. The total error rates for the unrecruited signals (both smeared and not smeared) decreased with the full audibility simulation, and the amounts of the confusions changed somewhat as well, though the locations generally remained the same. Presumably, increasing the audibility of the stimuli would allow for better discrimination because more spectral cues would be available, even if they were distorted from smearing.

It is interesting to note that the total error rates of the recruited signals did not decrease very much with increased audibility. The recruitment contributed to the signal degradation both by changing the level of the signal and by introducing spectral smearing, particularly at high frequencies. In real hearing-impaired subjects, using a compression hearing aid could counteract the level-changing effects of recruitment, but will not affect the frequency smearing effects. It would be interesting to compare the relative contributions of these effects of loudness recruitment, by simulating level changes without the frequency-smearing effects and comparing that to the standard recruitment algorithm preceded by a compression stage.
Finally, the relative performance of the Moore and ter Keurs smearing methods (with no recruitment) with reduced and full audibility merits discussion. In the reduced audibility case, the Moore method results fit the real subjects' confusion matrices better than the ter Keurs method results, but in the full-audibility case this difference disappeared. This could mean several things. It could be that the confusions by both the Moore and ter Keurs methods were so different than the real subjects' confusion matrices that it was by chance that the Moore method fit better in the reduced audibility simulation, and that neither really fits well. This is definitely a possibility given the large values of the $\chi^2$ measures of goodness of fit. A second possibility is that the Moore method produces better results only if there are minimal spectral cues. This could be true even if it is the case that the reduced audibility simulation is terribly inaccurate, and that in the real confusion matrix experiment many more frequencies were audible than in the simulation. If the reduced audibility simulation is inaccurate, and the real subjects heard most of the frequencies, this would imply that the Moore method is not better in general, but only for the situation in which the frequency cues are minimal.

4.4 Potential Perceptual Models in Humans

It was hypothesized that the primary inaccuracy in this simulation that caused relatively poor results was the fact that the the $d'$ discrimination model was different from how real human listeners do perception. The question remains as to what would be a better model of human perception, and this section addresses that question.

In the example given of two pairs of spectra that had the same $d'$ value but varied in how hard it would be for humans to discriminate between them (figure 4-2), the primary difference between spectral pairs was the presence or absence of distinct spectral features. In Case A, the spectra would be easy for humans to discriminate because they contained sharp spectral peaks and notches. In Case B, the spectra contained no distinct features, only a difference in level between the high and low
frequencies.

One possible perceptual model that would account for the ability of spectral features to help in signal discrimination is as follows. First, it is assumed that humans do discrimination based on the heights and frequencies of the formants. “Model” vowels (that would correspond to the response centers in this simulation) would be known to have formants of certain heights and at certain frequencies relative to one another. Discrimination of a sample unknown vowel would proceed by first identifying the formants present in the sample vowel’s spectrum, via the spectral peaks relative to spectral notches. The frequencies of these formants would be noted, and then the heights of the formants would be computed relative to an average vowel spectrum. A vector of cues would then be formed of the relative distances between the formants, and the heights of the formants. Finally, a $d'$ discrimination algorithm would be used on this cue vector, instead of on the original spectrum, and from this point the perceptual algorithm would proceed as before.
Figure 4-1: Average vowel lengths for the different vowels. The error bars represent one standard deviation of the vowel length. The averages were nearly identical for male and female speakers.
Either of Stimulus Sets A or B results in perceptual maps like those below, which have different results for the simulation and human listeners.

Figure 4-2: Examples of spectra. In both cases, the $d'$ discrimination algorithm will produce equal $d'$ distances between all stimuli. However, a real human listener could easily distinguish that the third spectrum was much different than the other two. In Stimulus Set A, the spectrum heights are A, 2A, and A, for spectra 1, 2, and 3, respectively. The widths of the spectra are such that the total area under all of the spectra is the same.
Chapter 5

Conclusions

5.1 Summary of Findings

Primarily, it was found that the $d'$ algorithm used in this simulation used spectral cues differently than how real humans do in perception. Because of this, confusions produced by the simulation tended to not occur in the same locations as the real subjects' confusions. This was evident because the unsmeared signals were judged to have confusions generally in the same locations as the confusions from the smeared signals.

Due to these factors, it was hard to judge the accuracy of the smearing models. In the case of limited audibility of the spectrum, such that just low-frequency cues are available, the Moore model did better than the ter Keurs model in terms of predicting the confusions made by the hearing-impaired listeners, but in both cases the fit was not very good.

Even so, increasing error rates with increasing smearing bandwidths occurred with the Moore and ter Keurs methods, both with and without recruitment. This shows that to some degree the smearing algorithms were acting correctly. Also, recruitment caused much-increased error rates, due to the spectral smearing it caused as well as the spectral level changes, which also is consistent with its acting correctly. In general, the error rates caused by the simulation seemed to be in support of the models
operating correctly, with the exception of the Moore model without recruitment, and the unsmeared stimuli.

We conclude that this method of simulation has potential to be very accurate, provided a good discrimination algorithm exists. Given the current discrimination algorithm, it is not possible to form strong conclusions about the relative accuracy of the Moore and ter Keurs models.

5.2 Relevance to Hearing Aid Signal Processing

Assuming that the model of loudness recruitment used in this simulation is correct, it was determined that recruitment causes spectral smearing as well as causing level changes in the spectrum. This is significant because it sets a limit on how well compression hearing aids could correct for loudness recruitment. Compression hearing aids could only correct for level changes, not frequency smearing, and would could possibly introduce additional spectral changes, as well. This limit could be further studied by simulations of recruitment preceded by a compression algorithm.

Also, this simulation seems to imply that there is a benefit to increasing the audibility of all frequencies, even if the ear has reduced frequency resolution at those frequencies. This conclusion is implied from the fact that in the simulation, reduced error rates were found in the simulation in which all frequencies were considered audible. This conclusion, however, is somewhat contingent on the accuracy of the discrimination model.

5.3 Recommendations for Future Work

There is much potential in having a simulation of hearing loss that does not require stimuli to be presented to and classified by normal-hearing subjects, because any stimuli played to them will necessarily be single waveforms, and they will be subjected to sharp frequency and temporal analysis by the subjects’ normal ears. Simulations
of hearing loss must necessarily operate on the different frequency bands of a signal, since in impaired ears the signal degradation occurs during the process of its being analyzed by the cochlea. However, if the results of processing in different frequency bands are added together again to form a single waveform for presentation to normal-hearing subjects, many of the effects of hearing loss will be undone.

In contrast, having multiple frequency bands as the output of a hearing loss simulator could produce quite accurate results, if the model is correct. However, some way of determining how humans would perceive the result of the processing is needed. So, future work on determining an accurate model of human perception could be quite beneficial. The results of such work could have application to speech recognition systems as well as to hearing aid research.

Once a good model of human perception is achieved, or even with the models currently available (assuming that some other way of analyzing the results is possible), work could focus on creating improved simulations of hearing loss that would feed into these perceptual models. One possibility for an improved model is described in Appendix D. This model attempts to accurately simulate all of the cochlear nonlinearities, so effects such as two-tone suppression could be seen. Additionally, it simulates the linear response of a location on the basilar membrane to low-frequency tones.

Finally, one other application for a model of human perception would be to determine to what extent the spectral analysis by a normal ear undoes the spectral smearing created by the hearing loss models currently employed that create output waveforms for presentation to normal-hearing listeners. Using a perceptual model, the limits on how much of different aspects of hearing loss could be simulated by presenting smeared sounds to normal-hearing listeners could be determined.
Appendix A

Audiograms and Confusion Matrices for the Real Subjects

Following are the audiograms of the subjects that participated in the confusion matrix experiment. The amounts of hearing loss were used to tailor the simulation to try achieve similar results to the subjects.

Figure A-1: Audiogram for subject PW.
Figure A-2: Audiograms for subjects JB and JO.
Figure A-3: Audiograms for subjects RC and MG.
Next are tables of the real subjects' actual confusion matrices. The confusion matrices listed here had the bias removed and were reconstructed using algorithm “a” in 4 dimensions by Dr. Louis Braida at MIT.

<table>
<thead>
<tr>
<th></th>
<th>AH</th>
<th>EE</th>
<th>EH</th>
<th>II</th>
<th>OO</th>
<th>UH</th>
</tr>
</thead>
<tbody>
<tr>
<td>AH</td>
<td>82</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>EE</td>
<td>1</td>
<td>98</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EH</td>
<td>0</td>
<td>0</td>
<td>98</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>II</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>89</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>OO</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>93</td>
<td>1</td>
</tr>
<tr>
<td>UH</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>17</td>
<td>55</td>
<td>14</td>
</tr>
</tbody>
</table>

Table A.1: Confusion matrix for subject JB.

<table>
<thead>
<tr>
<th></th>
<th>AH</th>
<th>EE</th>
<th>EH</th>
<th>II</th>
<th>OO</th>
<th>UH</th>
</tr>
</thead>
<tbody>
<tr>
<td>AH</td>
<td>88</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>EE</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EH</td>
<td>0</td>
<td>0</td>
<td>98</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>II</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>89</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>OO</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>88</td>
<td>4</td>
</tr>
<tr>
<td>UH</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>85</td>
</tr>
</tbody>
</table>

Table A.2: Confusion matrix for subject JO.

<table>
<thead>
<tr>
<th></th>
<th>AH</th>
<th>EE</th>
<th>EH</th>
<th>II</th>
<th>OO</th>
<th>UH</th>
</tr>
</thead>
<tbody>
<tr>
<td>AH</td>
<td>95</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>EE</td>
<td>0</td>
<td>93</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>EH</td>
<td>0</td>
<td>0</td>
<td>97</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>II</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>98</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OO</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>97</td>
<td>1</td>
</tr>
<tr>
<td>UH</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>51</td>
<td>46</td>
</tr>
</tbody>
</table>

Table A.3: Confusion matrix for subject MG.
Table A.4: Confusion matrix for subject PW.

<table>
<thead>
<tr>
<th></th>
<th>AH</th>
<th>EE</th>
<th>EH</th>
<th>II</th>
<th>OO</th>
<th>UH</th>
</tr>
</thead>
<tbody>
<tr>
<td>AH</td>
<td>89</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>EE</td>
<td>2</td>
<td>97</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EH</td>
<td>2</td>
<td>0</td>
<td>84</td>
<td>13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>II</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>88</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>OO</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>68</td>
<td>31</td>
</tr>
<tr>
<td>UH</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>39</td>
<td>60</td>
</tr>
</tbody>
</table>

Table A.5: Confusion matrix for subject RC.

<table>
<thead>
<tr>
<th></th>
<th>AH</th>
<th>EE</th>
<th>EH</th>
<th>II</th>
<th>OO</th>
<th>UH</th>
</tr>
</thead>
<tbody>
<tr>
<td>AH</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EE</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EH</td>
<td>0</td>
<td>0</td>
<td>93</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>II</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>92</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>OO</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>3</td>
<td>84</td>
<td>0</td>
</tr>
<tr>
<td>UH</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>92</td>
<td>0</td>
</tr>
</tbody>
</table>
Appendix B

Tables of Computed Confusion Matrices

The “standard amounts of smearing” used were: a widening factor of 3 for the Moore method, 0.75 octaves for the ter Keurs method, and 28 ms for the Hou method. So, for example, the Moore method in the Smearing Multiplier = 1.5 column simulates auditory filters widened by a factor of $1.5 \times 3 = 4.5$. 
Table B.1: Confusion matrices for the simulation done with subject JB’s parameters. From top to bottom, for each row the smearing conditions were: No smearing, Moore sm. only, ter Keurs sm. only, Hou sm. only. The amount of smearing in each column was the “Smearing Multiplier” times a 'standard amount of smearing' (see beginning of the appendix).
Table B.2: Confusion matrices for the simulation done with subject JB’s parameters. From top to bottom, for each row the smearing conditions were: Moore sm. then recruitment, ter Keurs sm. then recruitment, Hou sm. then recruitment, and just recruitment. The amount of smearing in each column was the “Smearing Multiplier” times a ‘standard amount of smearing’ (see beginning of the appendix).
Table B.3: Confusion matrices for the simulation done with subject JO’s parameters. From top to bottom, for each row the smearing conditions were: No smearing, Moore sm. only, ter Keurs sm. only, Hou sm. only. The amount of smearing in each column was the “Smearing Multiplier” times a 'standard amount of smearing' (see beginning of the appendix).
Table B.4: Confusion matrices for the simulation done with subject JO’s parameters. From top to bottom, for each row the smearing conditions were: Moore sm. then recruitment, ter Keurs sm. then recruitment, Hou sm. then recruitment, and just recruitment. The amount of smearing in each column was the “Smearing Multiplier” times a ‘standard amount of smearing’ (see beginning of the appendix).
Table B.5: Confusion matrices for the simulation done with subject MG's parameters. From top to bottom, for each row the smearing conditions were: No smearing, Moore sm. only, ter Keurs sm. only, Hou sm. only. The amount of smearing in each column was the “Smearing Multiplier” times a ‘standard amount of smearing’ (see beginning of the appendix).
<table>
<thead>
<tr>
<th></th>
<th>S.M. = 1.0</th>
<th></th>
<th>S.M. = 1.5</th>
<th></th>
<th>S.M. = 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AH EE EH II OO UH</td>
<td>AH EE EH II OO UH</td>
<td>AH EE EH II OO UH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AH</td>
<td>99 0 1 0 0 0 0</td>
<td>97 0 2 0 0 1</td>
<td>91 0 7 0 0 0 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE</td>
<td>88 0 1 11 0 0</td>
<td>84 0 1 14 0</td>
<td>81 0 3 16 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EH</td>
<td>5 1 80 3 0 11</td>
<td>8 1 75 4 0 12</td>
<td>11 1 72 6 0 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>5 0 5 69 17 9</td>
<td>1 6 56 15 23</td>
<td>3 9 50 13 25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OO</td>
<td>14 0 10 75 1</td>
<td>16 0 8 75 1</td>
<td>17 0 8 71 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UH</td>
<td>1 0 8 16 3 72</td>
<td>1 0 4 25 6 64</td>
<td>2 0 6 21 12 59</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>S.M. = 1.0</th>
<th></th>
<th>S.M. = 1.5</th>
<th></th>
<th>S.M. = 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AH EE EH II OO UH</td>
<td>AH EE EH II OO UH</td>
<td>AH EE EH II OO UH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AH</td>
<td>91 0 5 0 0 0 3</td>
<td>89 0 7 0 0 3</td>
<td>87 0 11 1 0 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE</td>
<td>85 1 3 11 0 0</td>
<td>84 1 2 13 0</td>
<td>82 1 2 15 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EH</td>
<td>5 1 62 11 1 21</td>
<td>9 1 52 16 2 20</td>
<td>15 1 48 15 5 18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>0 2 8 53 26 11</td>
<td>3 9 42 29 17</td>
<td>7 10 39 22 21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OO</td>
<td>20 0 17 60 3</td>
<td>17 1 21 58 3</td>
<td>17 0 20 59 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UH</td>
<td>3 0 19 20 2 56</td>
<td>3 0 18 23 1 55</td>
<td>3 0 20 22 4 51</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>S.M. = 1.0</th>
<th></th>
<th>S.M. = 1.5</th>
<th></th>
<th>S.M. = 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AH EE EH II OO UH</td>
<td>AH EE EH II OO UH</td>
<td>AH EE EH II OO UH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AH</td>
<td>100 0 0 0 0 0</td>
<td>100 0 0 0 0 0</td>
<td>100 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE</td>
<td>79 0 2 19 0 0</td>
<td>79 0 1 19 0</td>
<td>79 0 1 19 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EH</td>
<td>1 83 5 0 12</td>
<td>1 83 5 0 12</td>
<td>1 83 4 0 12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>0 0 7 77 12 4</td>
<td>0 0 7 77 12 4</td>
<td>0 0 7 76 12 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OO</td>
<td>18 0 5 75 2</td>
<td>19 0 5 75 2</td>
<td>19 0 5 75 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UH</td>
<td>1 0 3 15 3 78</td>
<td>1 0 3 15 3 78</td>
<td>1 0 3 15 3 77</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>S.M. = 1.0</th>
<th></th>
<th>S.M. = 1.5</th>
<th></th>
<th>S.M. = 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AH EE EH II OO UH</td>
<td>AH EE EH II OO UH</td>
<td>AH EE EH II OO UH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AH</td>
<td>100 0 0 0 0 0</td>
<td>100 0 0 0 0 0</td>
<td>100 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE</td>
<td>86 0 2 12 0</td>
<td>86 0 2 12 0</td>
<td>86 0 2 12 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EH</td>
<td>1 89 4 0 7</td>
<td>1 89 4 0 7</td>
<td>1 89 4 0 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>0 0 6 80 12 3</td>
<td>0 0 6 80 12 3</td>
<td>0 0 6 80 12 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OO</td>
<td>15 0 4 78 3</td>
<td>15 0 4 78 3</td>
<td>15 0 4 78 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UH</td>
<td>1 0 2 12 4 82</td>
<td>1 0 2 12 4 82</td>
<td>1 0 2 12 4 82</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table B.6: Confusion matrices for the simulation done with subject MG's parameters. From top to bottom, for each row the smearing conditions were: Moore sm. then recruitment, ter Keurs sm. then recruitment, Hou sm. then recruitment, and just recruitment. The amount of smearing in each column was the "Smearing Mulitplier" times a 'standard amount of smearing' (see beginning of the appendix).
Table B.7: Confusion matrices for the simulation done with subject RC’s parameters. From top to bottom, for each row the smearing conditions were: No smearing, Moore sm. only, ter Keurs sm. only, Hou sm. only. The amount of smearing in each column was the “Smearing Multiplier” times a ‘standard amount of smearing’ (see beginning of the appendix).
Table B.8: Confusion matrices for the simulation done with subject RC's parameters. From top to bottom, for each row the smearing conditions were: Moore sm. then recruitment, ter Keurs sm. then recruitment, Hou sm. then recruitment, and just recruitment. The amount of smearing in each column was the “Smearing Multiplier” times a ‘standard amount of smearing’ (see beginning of the appendix).
<table>
<thead>
<tr>
<th></th>
<th>S.M. = 1.0</th>
<th></th>
<th>S.M. = 1.5</th>
<th></th>
<th>S.M. = 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AH EE EH</td>
<td>II</td>
<td>OO UH</td>
<td>AH EE EH</td>
<td>II</td>
</tr>
<tr>
<td>AH</td>
<td>100 0 0 0 0</td>
<td>0</td>
<td>0 0 0 0</td>
<td>100 0 0 0</td>
<td>0</td>
</tr>
<tr>
<td>EE</td>
<td>0 91 0 0 9 0</td>
<td>0 91 0</td>
<td>0 0 9 0</td>
<td>0 91 0 0</td>
<td>9 0 0 9 0</td>
</tr>
<tr>
<td>EH</td>
<td>1 0 93 3 0 3</td>
<td>1 0 93</td>
<td>3 0 3 0</td>
<td>1 0 93 3</td>
<td>0 3 0 3 0</td>
</tr>
<tr>
<td>II</td>
<td>0 1 2 82 14 1</td>
<td>0 1 2</td>
<td>82 14 1 1</td>
<td>0 1 2 82</td>
<td>1 14 1 11</td>
</tr>
<tr>
<td>OO</td>
<td>0 8 0 8 80 3</td>
<td>0 8 0</td>
<td>8 80 3 0 3</td>
<td>0 8 0 8 0</td>
<td>8 80 3 0</td>
</tr>
<tr>
<td>UH</td>
<td>1 0 11 7 1 80</td>
<td>1 0 11</td>
<td>7 1 80 3</td>
<td>1 0 11 7</td>
<td>1 80 3 0</td>
</tr>
<tr>
<td>AH</td>
<td>99 0 1 0 0 0</td>
<td>95 0 5</td>
<td>0 0 0 0 0</td>
<td>91 0 8 0</td>
<td>0 0 0 0</td>
</tr>
<tr>
<td>EE</td>
<td>0 91 0 0 9 0</td>
<td>0 81 0</td>
<td>1 18 0 0 0</td>
<td>0 76 0</td>
<td>3 21 0 0</td>
</tr>
<tr>
<td>EH</td>
<td>3 1 81 2 0 14</td>
<td>6 1 75</td>
<td>3 0 16 0</td>
<td>8 0 68</td>
<td>5 0 19 0</td>
</tr>
<tr>
<td>II</td>
<td>0 0 3 71 18 8</td>
<td>0 1 3</td>
<td>54 20 22 0</td>
<td>0 3 8</td>
<td>48 22 19</td>
</tr>
<tr>
<td>OO</td>
<td>0 10 0 13 77 0</td>
<td>0 20 0</td>
<td>15 64 1 0</td>
<td>0 25 0</td>
<td>18 55 2</td>
</tr>
<tr>
<td>UH</td>
<td>1 0 11 15 2 72</td>
<td>1 0 18</td>
<td>26 6 49</td>
<td>2 0</td>
<td>20 28 8 42</td>
</tr>
<tr>
<td>AH</td>
<td>92 0 6 0 0 1</td>
<td>90 0 9</td>
<td>0 0 0 1 1</td>
<td>88 0 10</td>
<td>1 0 2 0</td>
</tr>
<tr>
<td>EE</td>
<td>0 86 0 2 12 0</td>
<td>0 82 0</td>
<td>1 15 0 0 0</td>
<td>0 80 0</td>
<td>1 18 0 0</td>
</tr>
<tr>
<td>EH</td>
<td>6 1 62 11 1 19</td>
<td>8 1 59</td>
<td>11 1 21 0</td>
<td>10 0 59</td>
<td>9 1 21 0</td>
</tr>
<tr>
<td>II</td>
<td>0 5 7 58 24 7</td>
<td>0 4 7</td>
<td>49 23 18</td>
<td>0 4</td>
<td>12 43 20</td>
</tr>
<tr>
<td>OO</td>
<td>0 15 1 19 60 5</td>
<td>0 15 1</td>
<td>21 59 4</td>
<td>0 17</td>
<td>2 20 57 4</td>
</tr>
<tr>
<td>UH</td>
<td>3 0 29 15 1 52</td>
<td>2 0</td>
<td>28 18 1 50</td>
<td>2 0</td>
<td>31 18 2 48</td>
</tr>
</tbody>
</table>

Table B.9: Confusion matrices for the simulation done with subject PW's parameters. From top to bottom, for each row the smearing conditions were: No smearing, Moore sm. only, ter Keurs sm! only, Hou sm. only. The amount of smearing in each column was the “Smearing Multiplier” times a 'standard amount of smearing' (see beginning of the appendix).
Table B.10: Confusion matrices for the simulation done with subject PW's parameters. From top to bottom, for each row the smearing conditions were: Moore sm. then recruitment, ter Keurs sm. then recruitment, Hou sm. then recruitment, and just recruitment. The amount of smearing in each column was the “Smearing Multiplier” times a 'standard amount of smearing’ (see beginning of the appendix).
Appendix C

Plots of $\chi^2$ Values

This appendix contains plots showing how well the simulated confusion matrices match the real subjects’ confusion matrices. To determine how well the confusion matrices match, a $\chi^2$ measure and a similar measure ($(O - E)^2$) were employed. The natural log of each is plotted in the top and bottom parts of each figure, respectively. Each plot compares one simulated confusion matrix to the real confusion matrices from all of the subjects. The title of each plot specifies the subject whose audiogram was used to generate the simulated confusion matrix, and the labels at the bottom of each column specify the real confusion matrix that it is being compared to. Each sub-column (indicated by the dotted lines) corresponds to one of the eight processing conditions. A label at the top of the sub-column indicates the processing condition for that column. The top symbol stands for the type of smearing, if any: ‘-’ = no smearing, ‘M’ = Moore smearing, ‘K’ = ter Keurs smearing, and ‘H’ = Hou smearing. The bottom symbol stands for whether the smearing was followed by recruitment or not: ‘-’ = no recruitment, and ‘R’ = recruitment. In the title, ‘S.M.’ stands for “smearing multiplier,” which is a scale factor indicating the amount of smearing. The actual amount of smearing used to generate the stimuli were (3 times the auditory filter width for the Moore method, 0.75 octaves for the ter Keurs method, and 28 msec for the Hou method) times the smearing multiplier. So, for example, if a plot says S.M. = 2.0 then that means e.g. the Moore method used smearing of 6 times
the normal auditory filter width. The first 15 plots in this appendix correspond to the reduced audibility condition (i.e. the normal simulation), and the last 15 plots correspond to data generated assuming that all frequencies were audible. These latter 15 are labeled as such in their captions.
Figure C-1: Original simulation including reduced audibility.
Figure C-2: Original simulation including reduced audibility.
Figure C-3: Original simulation including reduced audibility.
Figure C-4: Original simulation including reduced audibility.
Figure C-5: Original simulation including reduced audibility.
Figure C-6: Original simulation including reduced audibility.
In(Chi-squared) measure for the simulation fitted to subject MG with S.M.=1.0

Processing condition and real subject compared to the simulation

In( (O-E)^2 ) measure for the simulation fitted to subject MG with S.M.=1.0

Processing condition and real subject compared to the simulation

Figure C-7: Original simulation including reduced audibility.
Figure C-8: Original simulation including reduced audibility.
In(Chi-squared) measure for the simulation fitted to subject MG with S.M.=2.0

Figure C-9: Original simulation including reduced audibility.
Figure C-10: Original simulation including reduced audibility.
In(Chi-squared) measure for the simulation fitted to subject PW with S.M.=1.5

Processing condition and real subject compared to the simulation

Figure C-11: Original simulation including reduced audibility.
In(Chi-squared) measure for the simulation fitted to subject PW with S.M.=2.0

Figure C-12: Original simulation including reduced audibility.
Figure C-13: Original simulation including reduced audibility.
Figure C-14: Original simulation including reduced audibility.
Figure C-15: Original simulation including reduced audibility.
Figure C-16: All frequencies audible condition.
Figure C-17: All frequencies audible condition.
Figure C-18: All frequencies audible condition.
Figure C-19: All frequencies audible condition.
Figure C-20: All frequencies audible condition.
Figure C-21: All frequencies audible condition.
In(Chi-squared) measure for the simulation fitted to subject MG with S.M.=1.0

Figure C-22: All frequencies audible condition.
Figure C-23: All frequencies audible condition.
Figure C-24: All frequencies audible condition.
Figure C-25: All frequencies audible condition.
In(Chi-squared) measure for the simulation fitted to subject PW with S.M.=1.5

Processing condition and real subject compared to the simulation

Figure C-26: All frequencies audible condition.
In(Chi-squared) measure for the simulation fitted to subject PW with S.M.=2.0

Figure C-27: All frequencies audible condition.
Figure C-28: All frequencies audible condition.
In(Chi-squared) measure for the simulation fitted to subject RC with S.M. = 1.5

Figure C-29: All frequencies audible condition.
Figure C-30: All frequencies audible condition.
Appendix D

Recommendations for an Improved Simulation

D.1 Rationale and Background

It has been found that a location on the basilar membrane (BM) has a transfer function that acts compressively to tones at its characteristic frequency (CF), while it responds linearly to tones an octave or more below its CF (Oxenham and Plack, 1997). The current simulation does not take this effect into account. Also, it is desired in general to have an improved simulation of hearing loss. This appendix describes one possible model that should address these concerns.

D.2 Overview of Model

The model proposed here is largely based on a model of the BM’s response to low-side suppression. The model is described in Cai and Geisler (1996), and it simulates the response of a single location on the BM to a low-frequency suppressor tone and a tone at the CF of the BM location that is suppressed. The model described here uses the same principles found in that model, but expands it to multiple locations on the BM. It generally simulates the pattern of BM motion found in the ear, to produce
a set of "channels." These channels correspond to the pattern of BM motion that
would be detected by a single inner hair cell in the cochlea. Or, the channels could
be thought of as the outputs of a set of filters similar to auditory filters. Although
it is desirable to have a simulation that can produce a single waveform that can be
played to normal-hearing people in order to simulate hearing loss, it is probably not
possible to achieve all of the effects of hearing loss if this is done (Moore et al., 1992).
However, this thesis made use of a perceptual model that used the outputs of a set
of auditory filters to do discrimination. This model is potentially quite powerful,
as it does not require a signal to be analyzed by a normal ear first, which would
tend to remove the simulated effects of hearing loss due to its good frequency and
time resolution. Due to these potential benefits, a model of simulating hearing loss
is described that has multiple outputs corresponding to the outputs of a filter bank.
The resulting signal could be analyzed by either the perceptual model used in this
thesis, or another model that takes similar inputs.

D.3 Details of Model

A diagram outlining the model is shown in figure D-1. The model works as
follows. First, a signal is filtered into a number of frequency bands by a series of
logarithmically-spaced filters, such that the filters are of increasing width. This step
is crudely intended to approximate the filtering that would occur on the basilar
membrane in the absence of active amplification and tuning. The entire model,
including this step, uses filtering of the entire signal on a continuous-time basis–
that is, it does not break the input signal up into frames as was done in the Moore
and ter Keurs methods in this thesis. This is to prevent the negative effects of the
overlap-add method, as explained in Chapter 2 of this thesis.

What follows is easiest to explain from the perspective of how to form a single
channel at the output. Each channel, as previously stated, corresponds to the pattern
of BM motion that would be detected by a single inner hair cell. This pattern is similar
to that seen at the outputs of auditory filters, if they are simulated, but it differs in some ways. The channels include low frequencies with a gain of 1, while conventional auditory filters exclude low frequencies. Also, the frequency selectivity in general of these filters is probably different from that of typical auditory filters.

To produce a channel, the outputs from each of the initial bandpass filters are put through a set of gains, then added together. This is to simulate the magnitudes of travelling waves on the BM as they pass by a particular location corresponding to the channel. A plot of a possible set of gains corresponding to a single channel is shown in the bottom right portion of the diagram. At the top of the plot are shown the contributions from low frequencies: since a location on the BM responds linearly to frequencies an octave or more below its CF, the low frequencies contribute with a gain of 1 to the current channel being examined. In the middle of the plot, it can be seen that frequencies around the frequency corresponding to the CF of the channel are amplified (the details of how much amplification is used will be described shortly). At the bottom of the plot, the frequencies higher than the channel CF make no contribution to that location on the BM. Once the contributions of all the input frequencies are added together, the result is a channel.

The particular gains used on each of the input frequencies are not fixed, however. It is known that the BM responds compressively to sounds within about 30-90 dB SPL, and so this must be taken into account in the model. How this is done is based on the Cai and Geisler (1996) model previously mentioned. The model simulates the effect of outer hair cells causing active amplification of the BM motion. This is done by placing the signal corresponding to the BM motion at the location being examined (i.e., the channel, in this model) through a nonlinear transfer function that represents the response of OHCs. This transfer function between the BM motion and the “OHC output affecting the gain” is sigmoidal and asymmetric, as shown in the middle left of the diagram. This output from the OHC transfer function is then put through a lowpass filter to simulate how high-Q filters cannot respond instantaneously to changes in gain. Finally, the output from this low-pass filter is used to select a set
of gains for the input frequencies contributing to the channels.

The particular set of gains chosen to use on the input frequencies will depend on the level of the channel. It is in this manner that the compression of the BM is taken into consideration. Also, the set of gains will change shape somewhat with level. This results from the effect known as “the upward spread of masking;” the change in gain curves with level can be seen in the plot at the bottom of the diagram. In the plot, at higher levels the curves amplify more and more low frequencies. The exact curves used for the gains are the responses of the BM to off-CF tones, which are still being experimentally determined. Also, it should be noted that the feedback system between the channels and the gains of the input frequencies could vary considerably, as long as the appropriate set of gains is selected for a given level of the channel.

Due to this model’s combining different frequencies nonlinearly to simulate a single channel, various nonlinear phenomena of the BM such as suppression should be accurately simulated at the outputs.

D.4 Simulating Hearing Loss

The model described thus far is intended to simulate a normal ear. To alter the model to produce the responses corresponding to an impaired ear, the only parameter that needs changing is the outer hair cell transfer function. Since the compressive action of the outer hair cells drives the BM nonlinearities, this compression simply must be removed to simulate hearing loss. So, the outer hair cell transfer function should be made more linear, which will cause the on-CF frequencies to be amplified less, and thereby effectively include more of the surrounding frequencies, causing frequency smearing. Also, the level of the channel will be decreased because less gain is being applied. This simulates the absolute threshold reduction found in impaired ears. The effect of loudness recruitment is also implied through this reduction in the compressive nature of the outer hair cells.
D.5 Possible Sources of Error

In the paper by Cai and Geisler (1996), they commented that the cutoff frequency of the lowpass filter in their model needed to be adjusted for different frequency inputs. It may be the case that in this model as well, a fixed cutoff value for the lowpass filter for a given channel will not produce correct gains for all possible sets of input frequencies. Also, the filtering at the input of this model was somewhat arbitrarily determined, and may not be completely accurate. Finally, it may prove to be difficult to find the correct sets of gain curves for various channel levels.
Figure D-1: Diagram of new hearing loss model.
References


