

# Fortschritt-Berichte VDI

**Reihe 8**

Mess-, Steuerungs-  
und Regelungstechnik

Dipl.-Ing. Marcus Holmberg,  
Malmö

**Nr. 1162**

## **Speech Encoding in the Human Auditory Periphery: Modeling and Quantitative Assessment by Means of Automatic Speech Recognition**

Berichte aus dem

Institut für  
Automatisierungstechnik  
der TU Darmstadt



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Speech Encoding in the Early Auditory System . . . . .	1
1.1.1	The Speech Signal . . . . .	2
1.1.2	Speech Encoding . . . . .	4
1.1.3	Rate-Place Code . . . . .	5
1.1.4	Fine-Structure Cues . . . . .	6
1.1.5	Periodicity Cues . . . . .	6
1.1.6	Envelope Cues . . . . .	7
1.2	Automatic Speech Recognition Tools . . . . .	8
1.2.1	ASR Front End . . . . .	9
1.2.2	ASR Back End . . . . .	10
1.2.3	Hidden Markov Models . . . . .	10
1.2.4	HMMs as a Tool for Speech Encoding Assessment . . . . .	13
1.3	Automatic Speech Recognition with Auditory Models as Front End . . . . .	14
1.4	Structure of the Thesis . . . . .	15
<b>I</b>	<b>Modeling</b>	<b>17</b>
<b>2</b>	<b>A Model of the Human Auditory Periphery</b>	<b>18</b>
2.1	Introduction . . . . .	19
2.2	Model Description . . . . .	21
2.2.1	Middle Ear and Basilar Membrane Model . . . . .	21
2.2.2	Inner Hair Cell Model and Synaptic Mechanisms . . . . .	26
2.2.3	Replication of Human Data . . . . .	28
2.3	Model Results . . . . .	32
2.3.1	Filter Shapes . . . . .	32
2.3.2	Two-Tone Stimuli . . . . .	40
2.3.3	Temporal Properties . . . . .	43
2.3.4	Speech Stimuli . . . . .	46

2.4	Discussion . . . . .	50
2.5	Conclusions . . . . .	53
<b>3</b>	<b>A Model of Cochlear Nucleus Octopus Neurons</b>	<b>54</b>
3.1	Introduction . . . . .	55
3.2	Modeling . . . . .	56
3.2.1	Octopus Neuron Model . . . . .	56
3.3	Model Results . . . . .	59
3.3.1	Membrane Properties of the Octopus Neuron Model . . . . .	59
3.3.2	Responses to Pure Tone Stimuli . . . . .	60
3.3.3	Reverse Correlation . . . . .	62
3.3.4	Periodic Stimuli . . . . .	63
3.4	Discussion . . . . .	67
3.5	Conclusions . . . . .	68
<b>II</b>	<b>Using Automatic Speech Recognition as a Tool to Quantify Speech Discriminability in Spike Trains</b>	<b>69</b>
<b>4</b>	<b>Rate-Place Coding</b>	<b>70</b>
4.1	Introduction . . . . .	71
4.2	Model and Interfacing . . . . .	73
4.2.1	Auditory Model . . . . .	73
4.2.2	ASR Interface . . . . .	74
4.2.3	Recognition Task and Recognizer Back End . . . . .	75
4.3	Results . . . . .	76
4.3.1	Speech Coding in the Auditory Nerve Model . . . . .	76
4.3.2	High- and Low Spontaneous Rate Fibers . . . . .	77
4.3.3	The Effect of Sound Level on Recognition Rates . . . . .	79
4.3.4	How Many Auditory Nerve Fibers Are Needed for Speech Recognition? . . . . .	80
4.3.5	Comparison with the Speech Intelligibility Index . . . . .	82
4.4	Discussion . . . . .	84
4.4.1	The Rate-Place Coding Principle . . . . .	84
4.4.2	Gap to Human Performance . . . . .	85
4.5	Conclusions . . . . .	86
<b>5</b>	<b>Interpeak Interval Histograms</b>	<b>87</b>
5.1	Introduction . . . . .	88

5.2	Model and Interfacing . . . . .	89
5.2.1	Interpeak Interval Histograms . . . . .	90
5.2.2	ASR Interface . . . . .	91
5.3	Results . . . . .	93
5.3.1	Speech Representation in the IPIH Code . . . . .	95
5.3.2	How Many Auditory Nerve Fibers Do We Need for IPIH Coding? . . . . .	95
5.3.3	Recognition Results in Background Noise . . . . .	98
5.3.4	The Effect of Sound Level on Recognition Results . . . . .	99
5.3.5	Results on the Full Alphabet . . . . .	99
5.4	Discussion . . . . .	101
5.4.1	Comparison To Previous ASR Studies . . . . .	102
5.4.2	Revisiting the Gap to Human Performance . . . . .	103
5.5	Conclusions . . . . .	104
<b>6</b>	<b>Augmenting the Rate Code by Cochlear Nucleus Octopus Neurons</b>	<b>105</b>
6.1	Introduction . . . . .	106
6.2	Model and Interfacing . . . . .	107
6.3	Results . . . . .	110
6.3.1	How Many Octopus Neurons Are Required for Speech Encoding? . . . . .	111
6.3.2	Normal Sound Levels . . . . .	112
6.3.3	The Effect of Speech Level on the Recognition Scores . . . . .	113
6.3.4	Comparison with the Speech Intelligibility Index . . . . .	114
6.4	Discussion . . . . .	116
6.4.1	Using Periodicity Information . . . . .	118
6.5	Summary and Conclusions . . . . .	119
<b>7</b>	<b>Automatic Speech Recognition with an Adaptation Model Motivated by Auditory Processing</b>	<b>120</b>
7.1	Introduction . . . . .	121
7.2	Adaptation Physiology and Detailed Modeling . . . . .	122
7.2.1	Synaptic Adaptation . . . . .	122
7.2.2	The Effect of Adaption in a Detailed Physiological Model . . . . .	122
7.3	A Simplified Adaptation Model for ASR . . . . .	125
7.4	Design of Automatic Speech Recognition Experiments . . . . .	125
7.4.1	Feature Extraction . . . . .	126
7.4.2	Recognition Tasks and Recognizer Back End . . . . .	127
7.5	ASR Results and Discussion . . . . .	129
7.5.1	The Effect of the Adaptation Time Constant . . . . .	129

7.5.2	AURORA 2 Feature Extraction Performance Comparison . . . . .	130
7.5.3	AURORA 3 Feature Extraction Performance Comparison . . . . .	130
7.6	Conclusion . . . . .	132
<b>8</b>	<b>Summary and Outlook</b>	<b>133</b>
8.1	Summary and Discussion of the Results . . . . .	133
8.2	Future Directions . . . . .	136
<b>A</b>	<b>Model Description of the Inner Hair Cell – Auditory Nerve Complex</b>	<b>139</b>
A.1	Inner Hair Cell . . . . .	139
A.2	Vesicle Pool Dynamics . . . . .	140
A.2.1	Quantal Pool Dynamics . . . . .	141
A.2.2	Continuous Pool Dynamics . . . . .	142
A.3	Spike Generation . . . . .	143
A.3.1	Quantal Vesicle Release . . . . .	144
A.3.2	Continuous “Vesicle” Release . . . . .	144
A.3.3	Simplified Spike Generation . . . . .	144
<b>B</b>	<b>Free Field to Ear Drum Transfer Function</b>	<b>148</b>
<b>C</b>	<b>Full Equations of the Octopus Neuron Model</b>	<b>150</b>
<b>D</b>	<b>Speech Intelligibility Index Calculations</b>	<b>153</b>
D.1	Speech Spectrum Levels . . . . .	153
D.2	Noise Spectrum Levels . . . . .	153
D.3	SII Calculation . . . . .	154
D.4	Transforming SII to Recognition Scores . . . . .	155
	<b>Bibliography</b>	<b>157</b>