

Contents

1	Introduction	1
1.1	Motivation	1
1.2	The Meaning of Phase	2
1.3	Dual Microphone Speech Processing, or Why Two Ears Are Better Than One	3
1.4	The Microphone From the 22nd Century - The Human Ear	5
1.5	Why Smart Computers Are Hard To Find	6
1.6	The Bigger Picture	7
1.7	Book Overview	7
2	Signal Processing Basics	9
2.1	Continuous and Discrete Time Signals	9
2.2	Continuous Time Fourier Transform	12
2.2.1	Useful Mathematical Identities	13
2.3	Sampling	17
2.4	Spectral Analysis of Discrete Time Signals	20
2.4.1	The Effect of Sampling on the Fourier Transform of a Signal	20
2.4.2	The Reconstruction Theorem	23
2.4.3	The Discrete Time Fourier Transform (DTFT)	26
2.4.4	Sampling in the Frequency Domain	28
2.4.5	The Discrete Fourier Transform (DFT)	33
2.5	Windowing	48
2.6	Delaying Discrete Time Signals by Non-Integer Amounts	58

3	Single-Microphone Speech Processing	63
3.1	Introduction	63
3.2	Background	63
3.3	The Role of Phase in Speech Enhancement	64
3.4	The Role of Phase in Speech Recognition	66
3.4.1	The Fundamentals of HMM Based ASR	66
3.4.2	Performance Overview of ASR Systems	71
3.5	Phase Estimation from Magnitude	76
3.5.1	Signal Estimation from the Modified STFT	76
3.5.2	Signal Estimation from the STFT Magnitude	77
3.6	Recent Developments in Phase Utilization	78
3.7	Summary	81
4	Human Hearing	83
4.1	Anatomy of the Ear	83
4.1.1	External Ear	83
4.1.2	Middle Ear	84
4.1.3	Inner Ear	84
4.2	Physiology of the Ear	84
4.2.1	Transmission of Sound Through the Middle Ear	84
4.2.2	Physiology of the Cochlea	85
4.2.3	Inner Ear Performs Super Fast Fourier Transform	86
4.2.4	The “Place” Principle	86
4.2.5	Action Potential and Determination of Loudness	86
4.2.6	Detection of the Change in the Loudness and the Power Law	88
4.2.7	Threshold for Hearing and Frequency Range of Hearing	88
4.3	Hearing in the Central Nervous System	89
4.3.1	Parallel Processing of Sound in the Cerebral Cortex	89
4.3.2	Importance of the Cerebral Cortex in Hearing	89
4.4	The Importance of Phase in Human Speech Processing	90
4.5	Experimental Setup	91
4.6	Experimental Results	92
4.7	Modeling the Effect of Phase	94
4.8	Speech Recognition Experiments Incorporating Phase Restoration	97

4.8.1	Experimental Setup	97
4.8.2	Experimental Results	97
4.9	Conclusions	98
5	Multi-Microphone Phase-Based Speech Processing	101
5.1	Introduction	101
5.1.1	Dual Microphone Sound Model	105
5.1.2	Frequency Dependent Nature of Phase Wrapping	106
5.2	Delay-and-sum Beamforming	109
5.2.1	Two-microphone Sum Beamforming	109
5.2.2	Multi-element Sum Beamforming	112
5.2.3	Steering the Array	114
5.3	Sound Localization Using a Delay-and-sum Beamformer	116
5.4	TDOA Based Sound Localization	116
5.4.1	TDOA Estimation	119
5.5	A Detailed Look at the Phase Error	122
5.6	The Relationship Between Phase-Error and SNR	124
5.7	Probabilistic constraints on the SNRs	127
5.8	Phase-Based Time-Varying Filters	131
5.9	Beamforming as a Phase-Error Filter	134
6	Concluding Remarks	137
6.1	Summary (i.e. things you would have known if you had read the book)	137
6.2	Directions for Future Research	138
6.3	Where Does it End?	139
6.4	Disclaimer	140
	<i>Bibliography</i>	141