Kommunikation und Kybernetik in Einzeldarstellungen Herausgegeben von H. Wolter und W. D. Keidel 3

Speech Analysis Synthesis and Perception

Second Edition

By

James L. Flanagan

Head, Acoustics Research Department Bell Laboratories Murray Hill, New Jersey

With 258 Figures



Springer-Verlag \cdot New York \cdot Heidelberg \cdot Berlin 1972

ISBN 0-387-05561-4 Springer-Verlag New York · Heidelberg · Berlin ISBN 3-540-05561-4 Springer-Verlag Berlin · Heidelberg · New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically those of translation, reprinting, re-use of illustrations, broadcasting, reproduction by photocopying machine or similar means, and storage in data banks.

Under § 54 of the German Copyright Law where copies are made for other than private use, a fee is payable to the publisher, the amount of the fee to be determined by agreement with the publisher.

payable to the publisher, the amount of the let to be determined of congress Catalog Card Number (© by Springer-Verlag Berlin · Heidelberg 1965 and 1972. Library of Congress Catalog Card Number 70-172068. Printed in Germany. The use of general descriptive names, trade names, trade marks, etc. in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Act, may accordingly be used freely by anyone.

Typesetting, printing and binding: Universitätsdruckerei H.Stürtz AG, Würzburg

Preface to the Second Edition

The first edition of this book has enjoyed a gratifying existence. Issued in 1965, it found its intended place as a research reference and as a graduate-level text. Research laboratories and universities reported broad use. Published reviews—some twenty-five in number—were universally kind. Subsequently the book was translated and published in Russian (Svyaz; Moscow, 1968) and Spanish (Gredos, S.A.; Madrid, 1972).

Copies of the first edition have been exhausted for several years, but demand for the material continues. At the behest of the publisher, and with the encouragement of numerous colleagues, a second edition was begun in 1970. The aim was to retain the original format, but to expand the content, especially in the areas of digital communications and computer techniques for speech signal processing. As before, the intended audience is the graduate-level engineer and physicist, but the psychophysicist, phonetician, speech scientist and linguist should find material of interest.

Preparation of the second edition could not have advanced except for discussions, suggestions and advice from many colleagues. In particular, professors and scientists who have used the book in their university lectures, both here and abroad, provided valuable comment about organization and addition of new material. Also, research colleagues, especially my associates in the Acoustics Research Department at Bell Laboratories, provided critical assessment of technical data and views about emphasis. To list individually all who influenced these factors would require inordinate space. Rather, I commend to you their many scientific contributions described among the following pages. Naturally, any shortcomings in exposition or interpretation rest solely with me.

The task of examining page proofs was shared, with notable enthusiasm, among several associates. I owe special thanks to Doctors L. R. RABINER, R. W. SCHAFER, N. S. JAYANT, A. E. ROSENBERG, J. L. HALL, R. C. LUMMIS, J. M. KELLY and J. R. HASKEW for this assistance. Further, I am indebted to my company, Bell Laboratories, for supporting the work and making its facilities available for typing and drafting. My

Preface

secretary, Mrs. B. MASAITIS, bore the brunt of this work and deserves special praise. As earlier, the efficient staff of Springer, through the organization of Dr. H. MAYER-KAUPP, shielded me from many details in actualizing the printed volume.

Finally, again, to my wife and sons I express warm thanks for their contribution of weekends which might have been spent otherwise.

Warren Township, New Jersey January 15, 1972 James Flanagan

Preface to the First Edition

This book has its origin in a letter. In November of 1959, the late Prof. Dr. WERNER MEYER-EPPLER wrote to me, asking if I would contribute to a series he was planning on Communication. His book "Grundlagen und Anwendungen der Informationstheorie" was to serve us the initial volume of the series.

After protracted consideration, I agreed to undertake the job provided it could be done outside my regular duties at the Bell Telephone Laboratories. Shortly afterwards, I received additional responsibilities in my research organization, and felt that I could not conveniently pursue the manuscript. Consequently, except for the preparation of a detailed outline, the writing was delayed for about a year and a half. In the interim, Professor MEYER-EPPLER suffered a fatal illness, and Professors H. WOLTER and W. D. KEIDEL assumed the editorial responsibilities for the book series.

The main body of this material was therefore written as a leisuretime project in the years 1962 and 1963. The complete draft of the manuscript was duplicated and circulated to colleagues in three parts during 1963. Valuable comments and criticisms were obtained, revisions made, and the manuscript submitted to the publisher in March of 1964. The mechanics of printing have filled the remaining time.

If the reader finds merit in the work, it will be owing in great measure to the people with whom I have had the good fortune to be associated. In earlier days at the M.I.T. Acoustics Laboratory, my association with Professor K. N. STEVENS, Dr. A. S. HOUSE, and Dr. J. M. HEINZ was a great priviledge. During this same time, and on two separate occasions, Dr. G. FANT was a guest researcher at the M.I.T. laboratory. Later, during a summer, I had the priviledge of working as a guest in Dr. FANT's laboratory in Stockholm. On all occasions I profited from his views and opinion.

In more recent times, my associates at Bell Laboratories have been a constant stimulus and encouragement. Beginning with Dr. J. R.

١V

PIERCE, under whose direction research in speech and hearing has taken on renewed vigor, Doctors E. E. DAVID, Jr., M. R. SCHROEDER, M.V. MATHEWS, J. L. KELLY, Jr., N. GUTTMAN, P. B. DENES, G. G. HARRIS, and many, many others have provided sage advice, valuable collaboration and a stimulating research atmosphere. I am certain that this collection of technical talent is duplicated at no other place in the world.

I am greatly in the debt of numerous colleagues for valuable criticism and comment of the draft material. Their appraisals have aided materially in the revisions. Besides several of those already named, Professor G. E. PETERSON and Dr. H. K. DUNN, and a number of their associates at the University of Michigan, provided a wealth of valuable suggestions. Professor OSAMU FUJIMURA of the University of Electro-Communications, Tokyo, supplied many penetrating remarks, particularly on points relating to vocal-tract acoustics. Dr. W. A. VAN BERGEIJK of Bell Laboratories reviewed Chapter IV in detail. Messrs. A. M. NOLL, J. L. SULLIVAN, and H. R. SILBIGER, also of the Laboratories, studied the entire manuscript and supplied numerous helpful comments.

It is with deep regret that I conclude this effort without the counsel of Professor MEYER-EPPLER. I sincerely hope that it fulfills his original concept of the volume. I wish to express my appreciation to Professor WOLTER and to Professor KEIDEL for their continued support during the preparation. Also, the many details could not have been surmounted without the help of Dr. H. MAYER-KAUPP of Springer.

Finally, to my wife and family I express my deep appreciation for their contribution of my time.

Warren Township, New Jersey July 29, 1964 JAMES FLANAGAN

Contents 1.4. Analysis-Synthesis Telephony: An Approach to Improved Efficiency . 9

	3.23. The Acoustic "C" \ldots \ldots \ldots \ldots \ldots \ldots	30
	3.24. The Acoustic "G"	31
	3.25. Summary of the Analogous Acoustic Elements	34
3.3.	The Radiation Load at the Mouth and Nostrils	36
3.4.	Spreading of Sound About the Head	38
3.5.	The Source for Voiced Sounds	41
	3.51. Glottal Excitation	41
	3.52. Glottal Impedance	43
	3.53. Small-Signal Equivalent Source for the Glottis	50
3.6.	The Source for Noise and Transient Excitation of the Tract	53
3.7.	Some Characteristics of Vocal Tract Transmission	58
	3.71. Effect of Radiation Load upon Mode Pattern	61
	3.72. Effect of Glottal Impedance upon Mode Pattern	63
	3.73. Effect of Cavity Wall Vibration	65
	3.74. Two-Tube Approximation of the Vocal Tract	69
	3.75. Excitation by Source Forward in Tract	72
	3.76. Effects of the Nasal Tract	77
	3.77. Four-Tube, Three-Parameter Approximation of Vowel Produc-	
	tion	80
	3.78. Multitube Approximations and Electrical Analogs of the Vocal	
	Tract	83
3.8.	Fundamentals of Speech and Hearing in Analysis-Synthesis Telephony	85

IV. The Ear and Hearing	86
4.1. Mechanism of the Ear	87
4.11. The Outer Ear	88
4.12. The Middle Ear	88
4.13. The Inner Ear	91
4.14. Mechanical-to-Neural Transduction	94
4.15. Neural Pathways in the Auditory System	100
4.2. Computational Models for Ear Function	108
4.21. Basilar Membrane Model	109
4.22. Middle Ear Transmission.	112
4.23. Combined Response of Middle Ear and Basilar Membrane	114
4.24. An Electrical Circuit for Simulating Basilar Memorane Displace-	117
Menter Simulation of Membrane Motion	110
4.25. Computer Simulation of Memorane Motion	173
4.3 Illustrative Relations between Subjective and Physiological Rehavior	125
4.31. Pitch Perception	127
4.32. Binaural Lateralization	130
4.33. Threshold Sensitivity	136
4.34. Auditory Processing of Complex Signals	139
V. Techniques for Speech Analysis	140
5.1. Spectral Analysis of Speech	141
5.11. Short-Time Frequency Analysis	142
5.12. Measurement of Short-Time Spectra	144
5.13. Choice of the Weighting Function, $h(t)$.	147
5.14. The Sound Spectrograph	150
5.15. Short-Time Correlation Functions and Power Spectra	100
5.10. Average Fower Spectra	162
5.2 Formant Analysis of Speech	163
5.21. Formant-Frequency Extraction	165
5.211. Axis-Crossing Measures of Formant Frequency	166
5.212. Spectral Moments	167
5.213. Spectrum Scanning and Peak-Picking Methods	167
5.214. Digital Computer Methods for Formant Extraction	170
5.22. Measurement of Formant Bandwidth	181
5.3. Analysis of Voice Pitch	184
5.4. Articulatory Analysis of the Vocal Mechanism	187
5.5. Automatic Recognition of Speech	192
5.6. Automatic Recognition and Verification of Speakers	199
VI. Speech Synthesis	204
6.1. Mechanical Speaking Machines: Historical Efforts	205
6.2. Electrical Methods for Speech Synthesis	210
6.21. Spectrum Reconstruction Techniques	211
6.22. "Terminal Analog" Synthesizers	214
6.221. Terminal Properties of the Vocal Tract	214
6.222. Cascade Type Synthesizers	223
6.223. Parallel Synthesizers	225
6.23. Transmission-Line Analogs of the Vocal System	229

 6.24. Excitation of Electrical Synthesizers 6.241. Simulation of the Glottal Wave 6.242. Simulation of Unvoiced Excitation 6.243. Models for Sound Generation in the Vocal Tract 6.25. Vocal Radiation Factors 6.26. Speech Synthesis by Computer Simulation 6.261. Digital Techniques for Formant Synthesis 6.262. Digital Techniques for Vocal Tract Simulation 	232 232 245 246 259 260 260 272
VII. Perception of Speech and Speech-Like Sounds	276
7.1. Differential vs. Absolute Discrimination.	278
Speech \ldots	279
7.21. Limens for Vowel Formant Frequencies	279
7.22. Limens for Formant Amplitude	280
7.23. Limens for Formant Bandwidth	280
7.24. Limens for Fundamental Frequency	281
7.25. Limens for Excitation Intensity	281
7.26. Limens for Glottal Zeros	281
7.27. Discriminability of Maxima and Minima in a Noise Spectrum	282
7.28. Other Close-Comparison Measures Related to Speech	284
7.29. Differential Discriminations in the Articulatory Domain	286
7.3. Absolute Discrimination of Speech and Speech-Like Sounds	288
7.31. Absolute Identification of Phonemes	288
7.32. Absolute Identification of Syllables	291
7.33. Effects of Learning and Linguistic Association in Absolute	
Identification of Speech-Like Signals	297
7.34. Influence of Linguistic Association upon Differential Discrim-	
inability	301
7.4. Effects of Context and Vocabulary upon Speech Perception	303
7.5. The Perceptual Units of Speech	305
7.51. Models of Speech Perception	308
7.6. Subjective Evaluation of Transmission Systems	310
7.61. Articulation Tests \ldots \ldots \ldots \ldots \ldots	310
7.62. Quality Tests	312
7.7. Calculating Intelligibility Scores from System Response and Noise	
Level: The Articulation Index	315
7.8. Supplementary Sensory Channels for Speech Perception	318
7.81. Visible Speech Translator	318
7.82. Lactile vocoder	319
7.85. Low Frequency Vocoder	320
VIII. Systems for Analysis-Synthesis Telephony	321
8.1. Channel Vocoders	323
8.11. Design Variations in Channel Vocoders	325
8.12. Multiplexing Channel Vocoders	326
8.121. Frequency-Space Multiplexing	326
8.122. Time-Division Multiplexing	327
8.123. Digital Transmission of Vocoder Signals	328
8.13. Vocoder Performance	329

IX

8.2. Reduced Redundancy Channel Vocoders	. 330
8.21. "Peak-Picker"	330
8.22. Linear Transformation of Channel Signals	. 331
8.23. Pattern-Matching Vocoders.	332
8.3. Voice-Excited Vocoders	333
8.31. Multiplexing and Digitalization	336
8.4. Correlation Vocoders	337
8.5. Formant Vocoders	339
8.51. Multiplexing and Digitalization of Formant Vocoders	344
8.52. Voice-Excited Formant Vocoders	347
8.6. Orthogonal Function Vocoders	348
8.61. Expansion of the Speech Waveform	349
8.62. Expansion of the Short-Time Amplitude Spectrum	352
8.63. Expansion of the Short-Time Autocorrelation Function	354
8.7. Homomorphic Vocoders	361
8.8. Maximum Likelihood Vocoders	363
8.9. Linear Prediction Vocoders	367
8 10 Articulatory Vocoders	370
8 11 Frequency-Dividing Vocoders	370
8 111 Vohanc	371
8 112 Analytic Rooter	372
8 113 Harmonic Compressor	377
8 114 Phase Vocoder	378
8 12 Time-Assignment Transmission of Speech	386
8.13. Predictive Coding of Speech	390
8 131 Predictive Quantizing: Differential Pulse Code Modulation	395
8 132 Adaptive Predictive Coding	300
8 14 Delta Modulation	401
	.01
References.	406
Author Index	427
Subject Index	437

Speech Analysis Synthesis and Perception

Second Edition 1972