,Algorithms for Direction specific Processing of Sound Signals - the Realization of a binaural Cocktail-Party-Processor-System

- English Translation -

DISSERTATION
to obtain the Degree of a
Doctor-Engineer of the
Faculty of Electrical Engineering
at the Ruhr-University Bochum

of
HARALD SLATKY
Gelsenkirchen

Bochum 1992

Algorithmen zur richtungsselektiven Verarbeitung von Schallsignalen - die Realisierung eines binauralen Cocktail-Party-Prozessor-Systems

DISSERTATION
zur Erlangung des Grades eines
Doktor-Ingenieurs
der Fakultät für Elektrotechnik
an der Ruhr-Universität Bochum

von
HARALD SLATKY
Gelsenkirchen

Bochum 1992

Acknowledgement

The present thesis was developed as research associate at the institute "Lehrstuhl für allgemeine Elektrotechnik und Akustik (AEA)" at the Ruhr-University Bochum. I would like to express my sincere thanks to the holder of the chair, Prof. Dr.-Ing. J. Blauert, who made this thesis possible.

Many thanks to my former colleagues Dr.-Ing. Markus Bodden, Dr.-Ing. Werner Gaik, Priv.-Doz. Dr.-Ing. Herbert Hudde and Dr.-Ing. Siegbert Wolf and to all members of the staff and to all students at the institute AEA, for supporting me by their amicable cooperation and constructive feedback.

Last but not least I would like to thank Prof. Dr.-Ing. U. Heute for his friendly interest in this thesis and for agreeing to act as second examiner'

Contents

1. **Introduction** 1

2. **Psychoacoustical Background**
2.1. Recognition of Sound Directions 5
2.2. Processing of Signals of certain Directions 7
2.3. Properties and Performance of Binaural Signal Processing 8

3. **Auditory Experiments on the Localization of multiple Sound Sources**
3.1. Setup of the Auditory Experiments 9
3.2. Results of the Auditory Experiments 12
3.3. Theses about the Signal Processing of the Auditory System 22
3.4. Comparison between Auditory Experiments and Binaural Models 23

4. **Basics of binaural Signal Processing**
4.1. The Transmission Path from the Sound Source to the Ear 27
4.2. Binaural information at the Presence of multiple Sound Sources 31
4.3. Algorithms for binaural Processing of Ear Signals 38
4.4. Cross Correlation Models for a single Sound Source 39
4.5. The Cross Correlation Function for multiple Sound Sources 39
4.6. The complex Cross Product 44

5. **Algorithms for the Evaluation of interaural Phase Differences ("Phase-Difference-Cocktail-Party-Processor")**
5.1. Requirements for a Cocktail-Party-Processor 46
5.2. The interaural Cross Product 46
5.3. The Phase-Difference-Cocktail-Party-Processor 48
5.3.1. The interaural Cross Product at two Sound Sources 48
5.3.2. The Cocktail-Party-Processor-Algorithm 49
5.3.3. Signal Processing Examples 51
5.3.4. Single Sound Source with time variant Amplitude 55
5.4. Complex Sound Fields 57
5.4.1. The interaural Cross Product at arbitrary Sound Sources 57
5.4.2. Two Sound Sources with time variant Amplitudes 59
5.4.3. More than two Sound Sources with constant Amplitudes 61
5.4.4. Diffuse Sound Fields 61
5.5. Dominant Sources 64
5.6. Mapping of Sound Source Estimators onto a desired Direction 65
5.7. Conclusion 71

6. **Algorithms for the Evaluation of interaural Level Differences ("Level-Difference-Cocktail-Party-Processor")**
6.1. Problem Definition 74
6.2. Evaluation of Sound Source Attributes from the analysis of interaural Level Differences. 74
6.2.1. One Sound Source 74
6.2.2. Two Sound Sources 75
6.2.3. More than two Sound Sources 76
6.2.4. Diffuse Sound Field 77
6.2.5. Dominant Sources 79
6.3. Mapping of Sound Source Estimators on a desired Direction 81
6.4. Combining Estimators from interaural Time and Level Differences 83
6.5. Perspective: Combination of different spatial Analysis and Processing Methods 86

7. **A Signal Processing Framework for binaural Models**
7.1. Processing of the Input Signals 87
7.1.1. Critical Band Filters 87
7.1.2. Generation of the analytic Time Signals 89
7.1.3. Data Reduction 89
7.2. Processing of Output Signals 91
7.2.1. Requirements to a Re-Synthesis-Unit 91
7.2.2. Alignment of the Input Signals to the Signal Estimators 92
7.2.3. Reducing the Number of Output Channels 95
7.2.4. Generating the Time Function from Samples of the analytic Time Signal 95
7.3. Overall Presentation of the Cocktail-Party-Processor-Model 96
7.3.1. Model Structure 96
7.3.2. Capabilities of the Cocktail-Party-Processors 97

8. **Steering of Cocktail-Party-Processors**
8.1. Detection of the input Direction of Sound Sources: Hearing in Rooms 102
8.2. Dynamic Effects at the Detection of Directions: The Precedence-Effect 103
8.3. Modeling the Precedence-Effect by a binaural Cocktail-Party-Processor-Model 104
8.4. Consequences for the Steering of a Cocktail-Party-Processor 106
8.5. From Processor-control to a Precedence -Processor 107

9. **Summary and Perspective** 108

Appendix A: An Evaluation Method for Auditory Experiments 112
Appendix B: Critical Band Models 113
Appendix C: Simplified Free Field Outer Ear Transfer Functions 117
Appendix D: A flexible Filter Method in the Frequency Domain 118
Appendix E: Further Algorithms for Solving the Cocktail-Party-Processor-Problem 123
Appendix F: A Program Structure for complex Processes (Parallel Structures on sequential Computers) 125
Appendix G: Literature 129

Used Symbols

Indices

a	signal a
abt	sampling, sampled
b	signal b
c	signal c
d	diffuse field
f	free field transfer function
FG	critical band
G	weighting factor
HE	auditory event
i	summation index
k	summation index
korr	corrected value
l	left ear
m	transferred to the reference point center of the head
max	maximal value
min	minimal value
o	upper cut-off frequency of the critical band
p	summations index for sound sources
q	sound source (also summation index)
r	right ear
soll	value of the desired direction
SQ	sound source
u	lower cut-off frequency of the critical band
x	parameter of an arbitrary estimator x (x= a or b)
μ	mean value
σ	standard deviation
θ	input direction

Exponents

'	estimator
+	cross power density of a signal
*	conjugated complex

Formula Symbols

$a(t)$	time function of the sound signal a
$a_m(t)$	time function of the sound signal a at the reference point center of the head
$a_m'(t)$	estimator for the sound signal amplitude a at the reference point center of the head
$\underline{a}_m(t)$	analytic time function of the sound signal a at the reference point center of the head
$\underline{A}(f)$	Fourier-transform of the sound signal a
$\underline{A}_m(t)^2$	source vector (at the reference point center of the head)
$\underline{A}_m'(t)^2$	source estimator (at the reference point center of the head)
$\underline{A}_m(f,\tau)$	Fourier-transform of he sound signal a at the reference point center of the head
$\underline{A}_m^+(f,t)^2$	spectral cross power density of sound signal a(t) at the reference point center of the head
AntwB	range of answers of the auditory experiments (e.g.. ±90°)
$b(t)$	time function of the sound signal b
$b_m(t)$	time function of the sound signal b at the reference point center of the head
$b_m'(t)$	estimator for the sound signal amplitude b at the reference point center of the head
$\underline{b}_m(t)$	analytic time function of the sound signal b at the reference point center of the head
$\underline{B}(f)$	Fourier transform of the sound signal b
$\underline{B}_m(t)^2$	source vector (at the reference point center of the head)
$\underline{B}_m'(t)^2$	sound source estimator (at the reference point center of the head)
$\underline{B}_m(f,\tau)$	Fourier transform of the sound signal at the reference point center of the head
$\underline{B}_m^+(f,\tau)^2$	spectral cross power density of sound signal b(t) at the reference point center of the head
c_{schall}	sound velocity
d	microphone distance
$\underline{d}(t)$	interaural difference
$E_m'^2$	power density of the diffuse sound field
$\underline{e}_{m\theta}$	analytic time signal of mirror sources
f	frequency
f_{abt}	sampling rate
$\underline{f}_{korr}$	correction factor
f_m	center frequency of the critical band
f_{min}	lower limit of the considered frequency range
f_{max}	upper limit of the considered frequency range
f_o	upper cut-off frequency
f_{oi}	upper cut-off frequency of the critical band i
f_u	lower cut-off frequency
f_{ui}	lower cut-off frequency of the critical band i
$f()$	function of
F	lock-in range around a sound source, where auditory events are considered as correctly localized
$\mathcal{F}$	Fourier transform

$\mathcal{F}^{-1}$	inverse Fourier transform
FG	critical band
g	weighting factor
$\underline{G}_m'^2$	compensation signal for correction purposes
$h_l(t,\tau)$	outer ear impulse response of the left ear
$h_r(t,\tau)$	outer ear impulse response of the right ear
$\underline{H}_l(f,\tau)$	transfer function left ear - center of the head
$\underline{H}_{lf}(f,\tau)$	free field-outer ear -transfer function of the left ear
$\underline{H}_m(f,\tau)$	free field-transfer function head position - center of the head
$\underline{H}_{qk}(f,\tau)$	free field-transfer function sound source - head position
$\underline{H}_{ql}(f,\tau)$	transfer function sound source - left ear
$\underline{H}_{qr}(f,\tau)$	transfer function sound source - right ear
$\underline{H}_r(f,\tau)$	transfer function right ear - center of the head
$\underline{H}_{rf}(f,\tau)$	free field-outer ear -transfer function of the right ear
$\underline{H}_{rl}(f,\tau)$	interaural transfer function
HE	auditory event
$\underline{k}(t)$	cross product
l(t)	real time function of the left ear signal.
$\underline{l}(t)$	analytic time signal of the left ear signal.
$\underline{L}(f)$	Fourier transform of the left ear signal.
LG	localization rate: degree of correspondence between sound direction and auditory event
Lt	lateralization (-10=left, 0=center, 10=right)
n_{oi}	slew rate of the upper filter slope of the critical band i
n_{ui}	slew rate of the lower filter slope of the critical band i
M, N	total amount
N_{HE}	number of perceived auditory events in an auditory experiment
N_{SQ}	number of sound sources
r(t)	real time function of the right ear signal.
$\underline{r}(t)$	analytic time signal of the right ear signal.
$\underline{R}(f)$	Fourier transform of the right ear signal.
s	signal of the desired direction
s_m	signal of the desired direction at the reference point center of the head
$s_m'^2$	signal power estimator for the desired direction at the reference point center of the head
$\underline{S}_m'^2$	source vector estimator for the desired direction (reference point center of the head)
$S_{rl}(\tau)$	cross correlation function
$S_{rl}(\tau,t)$	sliding cross correlation function
$\underline{S}_{rl}(f,t)$	Fourier transform of the sliding cross correlation function
SE	sound event
t	time

T	time interval
T_μ	integration time for the evaluation of the statistical parameters
Vers	experiment (number)
VP	test person(number)
w_r	guess probability
$w(t)$	window function
$\underline{W}(f)$	Fourier transform of the window function
W_x	weighting function for the estimator x
x	arbitrary estimator (x = a or b)
$x_m'^2$	power of an arbitrary estimator at the reference point center of the head
$\underline{X}_m'^2$	arbitrary source estimator (reference point center of the head)
α	interaural damping
α'	estimator for the interaural damping
β	interaural phase
β'	estimator for the interaural phase
β_x'	interaural phase of the arbitrary source estimator $\underline{X}_m'^2$
Δf	(center-)frequency difference of auditory experiment signals
Δf_L	(center-)frequency difference, at which one sound source can be localized correctly.
Δf_{L2}	(center-)frequency difference, at which both sound sources can be localized correctly.
Δf_{FG}	bandwidth of a critical band
ΔL	interaural level difference
$\underline{\mu}$	complex mean value of the cross correlation function or of the cross product
μ_l	mean value of the power of the left ear signal
μ_r	mean value of the power of the right ear signal
Φ	signal phase
$\Delta\Phi_{ab}$	difference between the signal phases of the signals a and b
Ω	current angular velocity
$\underline{\sigma}$	complex standard deviation of the cross correlation function or of the cross product
σ_l	standard deviation of the power of the left ear signal
σ_r	standard deviation of the power of the right ear signal
τ	interaural time difference
τ'	additional interaural time difference (cross correlation function)
τ_L	normalized interaural time difference (±90° corresponds to ±625µs)
τ_{max}	maximal interaural time difference
τ_m	mean interaural time difference inside the considered frequency range
τ_o	mean runtime source-receiver
θ	angle of incidence