Computational Auditory Scene Analysis. Principles, Algorithms, and Applications

Description: How can we engineer systems capable of "cocktail party" listening?

Human listeners are able to perceptually segregate one sound source from an acoustic mixture, such as a single voice from a mixture of other voices and music at a busy cocktail party. How can we engineer "machine listening" systems that achieve this perceptual feat?

Albert Bregman's book Auditory Scene Analysis, published in 1990, drew an analogy between the perception of auditory scenes and visual scenes, and described a coherent framework for understanding the perceptual organization of sound. His account has stimulated much interest in computational studies of hearing. Such studies are motivated in part by the demand for practical sound separation systems, which have many applications including noise-robust automatic speech recognition, hearing prostheses, and automatic music transcription. This emerging field has become known as computational auditory scene analysis (CASA).

Computational Auditory Scene Analysis: Principles, Algorithms, and Applications provides a comprehensive and coherent account of the state of the art in CASA, in terms of the underlying principles, the algorithms and system architectures that are employed, and the potential applications of this exciting new technology. With a Foreword by Bregman, its chapters are written by leading researchers and cover a wide range of topics including:
- Estimation of multiple fundamental frequencies
- Feature-based and model-based approaches to CASA
- Sound separation based on spatial location
- Processing for reverberant environments
- Segregation of speech and musical signals
- Automatic speech recognition in noisy environments
- Neural and perceptual modeling of auditory organization

The text is written at a level that will be accessible to graduate students and researchers from related science and engineering disciplines. The extensive bibliography accompanying each chapter will also make this book a valuable reference source. A web site accompanying the text, company website, features software tools and sound demonstrations.

Contents:
- Foreword.
- Preface.
- Contributors.
- Acronyms.
  1.1 Human Auditory Scene Analysis.
    1.1.1 Structure and Function of the Auditory System.
    1.1.2 Perceptual Organization of Simple Stimuli.
    1.1.3 Perceptual Segregation of Speech from Other Sounds.
    1.1.4 Perceptual Mechanisms.
  1.2 Computational Auditory Scene Analysis (CASA).
1.2.1 What Is CASA?
1.2.2 What Is the Goal of CASA?
1.2.3 Why CASA?
1.3 Basics of CASA Systems.
1.3.1 System Architecture.
1.3.2 Cochleagram.
1.3.3 Correlogram.
1.3.4 Cross-Correlogram.
1.3.5 Time-Frequency Masks.
1.3.6 Resynthesis.
1.4 CASA Evaluation.
1.4.1 Evaluation Criteria.
1.4.2 Corpora.
1.5 Other Sound Separation Approaches.
1.6 A Brief History of CASA (Prior to 2000).
1.6.1 Monaural CASA Systems.
1.6.2 Binaural CASA Systems.
1.6.3 Neural CASA Models.
1.7 Conclusions

Acknowledgments.

References.

2. Multiple F0 Estimation (Alain de Cheveigné).
2.1 Introduction.
2.2 Signal Models.
2.3 Single-Voice F0 Estimation.
2.3.1 Spectral Approach.
2.3.2 Temporal Approach.
2.3.3 Spectrotemporal Approach.
2.4 Multiple-Voice F0 Estimation.
2.4.1 Spectral Approach.
2.4.2 Temporal Approach.
2.4.3 Spectrotemporal Approach.
2.5 Issues.
2.5.1 Spectral Resolution.
2.5.2 Temporal Resolution.
2.5.3 Spectrotemporal Resolution.
2.6 Other Sources of Information.
2.6.1 Temporal and Spectral Continuity.
2.6.2 Instrument Models.
2.6.3 Learning-Based Techniques.
2.7 Estimating the Number of Sources.
2.8 Evaluation.
2.9 Application Scenarios.
2.10 Conclusion.

Acknowledgments.

References.

3. Feature-Based Speech Segregation (DeLiang Wang).

3.1 Introduction.

3.2 Feature Extraction.
3.2.1 Pitch Detection.
3.2.2 Onset and Offset Detection.
3.2.3 Amplitude Modulation Extraction.
3.2.4 Frequency Modulation Detection.

3.3 Auditory Segmentation.
3.3.1 What Is the Goal of Auditory Segmentation?
3.3.2 Segmentation Based on Cross-Channel Correlation and Temporal Continuity.
3.3.3 Segmentation Based on Onset and Offset Analysis.

3.4 Simultaneous Grouping.
3.4.1 Voiced Speech Segregation.
3.4.2 Unvoiced Speech Segregation.

3.5 Sequential Grouping.
3.5.1 Spectrum-Based Sequential Grouping.
3.5.2 Pitch-Based Sequential Grouping.
3.5.3 Model-Based Sequential Grouping.

3.6 Discussion.

Acknowledgments.

References.

4. Model-Based Scene Analysis (Daniel P. W. Ellis).

4.1 Introduction.

4.2 Source Separation as Inference.

4.3 Hidden Markov Models.

4.4 Aspects of Model-Based Systems.

4.4.1 Constraints: Types and Representations.

4.4.2 Fitting Models.

4.4.3 Generating Output.

4.5 Discussion.

4.5.1 Unknown Interference.

4.5.2 Ambiguity and Adaptation.

4.5.3 Relations to Other Separation Approaches.

4.6 Conclusions.

References.


5.1 Introduction.

5.2 Physical and Physiological Mechanisms Underlying Auditory Localization.

5.2.1 Physical Cues.

5.2.2 Physiological Estimation of ITD and IID.

5.3 Spatial Perception of Single Sources.

5.3.1 Sensitivity to Differences in Interaural Time and Intensity.

5.3.2 Lateralization of Single Sources.

5.3.3 Localization of Single Sources.

5.3.4 The Precedence Effect.

5.4 Spatial Perception of Multiple Sources.

5.4.1 Localization of Multiple Sources.

5.4.2 Binaural Signal Detection.

5.5 Models of Binaural Perception.
5.5.1 Classical Models of Binaural Hearing.

5.5.2 Cross-Correlation-Based Models of Binaural Interaction.

5.5.3 Some Extensions to Cross-Correlation-Based Binaural Models.

5.6 Multisource Sound Localization.

5.6.1 Estimating Source Azimuth from Interaural Cross-Correlation.

5.6.2 Methods for Resolving Azimuth Ambiguity.

5.6.3 Localization of Moving Sources.

5.7 General Discussion.

Acknowledgments.

References.


6.1 Introduction.

6.2 Classical Beamforming Techniques.

6.2.1 Fixed Beamforming Techniques.

6.2.2 Adaptive Beamforming Techniques.

6.2.3 Independent Component Analysis Techniques.

6.2.4 Other Localization-Based Techniques.

6.3 Location-Based Grouping Using Interaural Time Difference Cue.

6.4 Location-Based Grouping Using Interaural Intensity Difference Cue.

6.5 Location-Based Grouping Using Multiple Binaural Cues.

6.6 Discussion and Conclusions.

Acknowledgments.

References.


7.1 Introduction.

7.2 Effects of Reverberation on Listeners.

7.2.1 Speech Perception.

7.2.2 Sound Localization.

7.2.3 Source Separation and Signal Detection.

7.2.4 Distance Perception.

7.2.5 Auditory Spatial Impression.
7.3 Effects of Reverberation on Machines.
7.4 Mechanisms Underlying Robustness to Reverberation in Human Listeners.
7.4.1 The Role of Slow Temporal Modulations in Speech Perception.
7.4.2 The Binaural Advantage.
7.4.3 The Precedence Effect.
7.4.4 Perceptual Compensation for Spectral Envelope Distortion.
7.5 Reverberation-Robust Acoustic Processing.
7.5.1 Dereverberation.
7.5.2 Reverberation-Robust Acoustic Features.
7.5.3 Reverberation Masking.
7.6 CASA and Reverberation.
7.6.1 Systems Based on Directional Filtering.
7.6.2 CASA for Robust ASR in Reverberant Conditions.
7.6.3 Systems that Use Multiple Cues.
7.7 Discussion and Conclusions.
Acknowledgments.
References.
8.1 Introduction.
8.2 Music Scene Description.
8.2.1 Music Scene Descriptions.
8.2.2 Difficulties Associated with Musical Audio Signals.
8.3 Estimating Melody and Bass Lines.
8.3.1 PreFEst-front-end: Forming the Observed Probability Density Functions.
8.3.2 PreFEst-core: Estimating the F0's Probability Density Function.
8.3.3 PreFEst-back-end: Sequential F0 Tracking by Multiple-Agent Architecture.
8.3.4 Other Methods.
8.4.1 Estimating Period and Phase.
8.4.2 Dealing with Ambiguity.
8.4.3 Using Musical Knowledge.
8.5 Estimating Chorus Sections and Repeated Sections.
8.5.1 Extracting Acoustic Features and Calculating Their Similarity.
8.5.2 Finding Repeated Sections.
8.5.3 Grouping Repeated Sections.
8.5.4 Detecting Modulated Repetition.
8.5.5 Selecting Chorus Sections.
8.5.6 Other Methods.
8.6 Discussion and Conclusions.
8.6.1 Importance.
8.6.2 Evaluation Issues.
8.6.3 Future Directions.
References.

9. Robust Automatic Speech Recognition (Jon Barker).
9.1 Introduction.
9.2 ASA and Speech Perception in Humans.
9.2.1 Speech Perception and Simultaneous Grouping.
9.2.2 Speech Perception and Sequential Grouping.
9.2.3 Speech Schemes.
9.2.4 Challenges to the ASA Account of Speech Perception.
9.2.5 Interim Summary.
9.3 Speech Recognition by Machine.
9.3.1 The Statistical Basis of ASR.
9.3.2 Traditional Approaches to Robust ASR.
9.3.3 CASA-Driven Approaches to ASR.
9.4 Primitive CASA and ASR.
9.4.1 Speech and Time-Frequency Masking.
9.4.2 The Missing-Data Approach to ASR.
9.4.3 Marginalization-Based Missing-Data ASR Systems.
9.4.4 Imputation-Based Missing-Data Solutions.
9.4.5 Estimating the Missing-Data Mask.
9.4.6 Difficulties with the Missing-Data Approach.
9.5 Model-Based CASA and ASR.
9.5.1 The Speech Fragment Decoding Framework.
9.5.2 Coupling Source Segregation and Recognition.
9.6 Discussion and Conclusions.
9.7 Concluding Remarks.

References.


10.1 Introduction.

10.2 The Neural Basis of Auditory Grouping.

10.2.1 Theoretical Solutions to the Binding Problem.

10.2.2 Empirical Results on Binding and ASA.

10.3 Models of Individual Neurons.

10.3.1 Relaxation Oscillators.

10.3.2 Spike Oscillators.

10.3.3 A Model of a Specific Auditory Neuron.

10.4 Models of Specific Perceptual Phenomena.

10.4.1 Perceptual Streaming of Tone Sequences.

10.4.2 Perceptual Segregation of Concurrent Vowels with Different F0s.

10.5 The Oscillatory Correlation Framework for CASA.

10.5.1 Speech Segregation Based on Oscillatory Correlation.

10.6 Schema-Driven Grouping.

10.7 Discussion.

10.7.1 Temporal or Spatial Coding of Auditory Grouping.

10.7.2 Physiological Support for Neural Time Delays.

10.7.3 Convergence of Psychological, Physiological, and Computational Approaches.

10.7.4 Neural Models as a Framework for CASA.

10.7.5 The Role of Attention.

10.7.6 Schema-Based Organization.

Acknowledgments.

References.

Index.
Order by Fax - using the form below

Order by Post - print the order form below and send to

Research and Markets,
Guinness Centre,
Taylors Lane,
Dublin 8,
Ireland.
Fax Order Form

To place an order via fax simply print this form, fill in the information below and fax the completed form to 646-607-1907 (from USA) or +353-1-481-1716 (from Rest of World). If you have any questions please visit http://www.researchandmarkets.com/contact/

Order Information
Please verify that the product information is correct.

<table>
<thead>
<tr>
<th>Product Name:</th>
<th>Computational Auditory Scene Analysis. Principles, Algorithms, and Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web Address:</td>
<td><a href="http://www.researchandmarkets.com/reports/2181380/">http://www.researchandmarkets.com/reports/2181380/</a></td>
</tr>
<tr>
<td>Office Code:</td>
<td>SCDVVBDQ</td>
</tr>
</tbody>
</table>

Product Format
Please select the product format and quantity you require:

**Quantity**

- Hard Copy (Hard Back): [ ] USD 128 + USD 28 Shipping/Handling

* Shipping/Handling is only charged once per order.

Contact Information
Please enter all the information below in **BLOCK CAPITALS**

<table>
<thead>
<tr>
<th>Title:</th>
<th>Mr [ ]</th>
<th>Mrs [ ]</th>
<th>Dr [ ]</th>
<th>Miss [ ]</th>
<th>Ms [ ]</th>
<th>Prof [ ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Name:</td>
<td></td>
<td></td>
<td></td>
<td>Last Name:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email Address:</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job Title:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organisation:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Address:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>City:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postal / Zip Code:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phone Number:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fax Number:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Please refrain from using free email accounts when ordering (e.g. Yahoo, Hotmail, AOL)
Payment Information

Please indicate the payment method you would like to use by selecting the appropriate box.

☐ Pay by credit card: You will receive an email with a link to a secure webpage to enter your credit card details.

☐ Pay by check: Please post the check, accompanied by this form, to:

Research and Markets,
Guinness Center,
Taylors Lane,
Dublin 8,
Ireland.

☐ Pay by wire transfer: Please transfer funds to:

<table>
<thead>
<tr>
<th>Description</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Account number</td>
<td>833 130 83</td>
</tr>
<tr>
<td>Sort code</td>
<td>98-53-30</td>
</tr>
<tr>
<td>Swift code</td>
<td>ULSBIE2D</td>
</tr>
<tr>
<td>IBAN number</td>
<td>IE78ULSB98533083313083</td>
</tr>
<tr>
<td>Bank Address</td>
<td>Ulster Bank, 27-35 Main Street, Blackrock, Co. Dublin, Ireland.</td>
</tr>
</tbody>
</table>

If you have a Marketing Code please enter it below:

Marketing Code:

Please note that by ordering from Research and Markets you are agreeing to our Terms and Conditions at http://www.researchandmarkets.com/info/terms.asp

Please fax this form to:

(646) 607-1907 or (646) 964-6609 - From USA
+353-1-481-1716 or +353-1-653-1571 - From Rest of World