ACOUSTICAL·SOCIETY·OF·AMERICA

1993-94 F. V. HUNT POSTDOCTORAL RESEARCH FELLOWSHIP

IN ACOUSTICS

GENERAL APPLICATION INFORMATION

1.   *ELIGIBILITY*.   To be eligible for the Fellowship a person shall have recently received his or her doctorate degree, or shall be receiving the degree in the Spring of 1993.

Qualification for the award requires completion of an Application Form and of Attachment A thereto. Attachment A provides for agreement by the institution where the research is to be conducted that proper facilities will be made available to the candidate, if selected. It is the responsibility of candidates to obtain the requisite approval from their chosen institution.

Each candidate shall be a member of the Acoustical Society of America prior to applying for the Hunt Fellowship.

2.   *CRITERIA FOR THE FELLOWSHIP*. The recipient of the Fellowship will be that individual who, through personal qualifications and a proposed research experience, is judged to exhibit the highest potential for performing research benefiting some aspect of the science of sound and promoting its usefulness to society.

3.   *ADMINISTRATION*.   The recipient of the 1993 award will be notified not later than 1 February 1993. The usual twelve-month tenure begins in the interval June--September; however, other specific arrangements will be considered. The amount of the 1993 award will be $27,500. Of this sum, $24,500 will be applied to stipend, and up to $3,000 will be reimbursable for travel, computer charges, publication page charges, material, and other costs.

4.   *APPLICATION DEADLINE*. The deadline for applications for the 1993 Fellowship Award is 1 September 1992. Applications postmarked after that date will not be accepted.

5.   *APPLICATION ADDRESS*. Applications should be forwarded to the Acoustical Society of America, 500 Sunnyside Blvd., Woodbury, New York 11797.

ACOUSTICAL·SOCIETY·OF·AMERICA

F. V. HUNT POSTDOCTORAL RESEARCH FELLOWSHIP IN ACOUSTICS
APPLICATION FORM

1. NAME OF CANDIDATE _____

   HOME ADDRESS _____

   Last   First   Middle

   _____

   _____ BIRTH DATE _____

   DAYTIME PHONE NUMBER _____

2. UNDERGRADUATE AND GRADUATE EDUCATION   CITIZEN OF _____

   Educational Institution   Department   Location   Degree & Date*

*If the Ph.D. is not yet awarded, place anticipated month and year in parentheses.

3. TITLE OF DOCTORAL THESIS

4. LOCATION, TOPIC AND TIME PERIOD OF PROPOSED RESEARCH

5. REFERENCES

   Two references from professional associates, supervisors or faculty are required. The applicant is responsible for soliciting these and ensuring that they are sent to the ASA Executive Director before the deadline, 1 September.

6. SUPPORTING INFORMATION   Please attach in the following order:
   a) Synopsis of Thesis
   b) Summary of Proposed Research (not to exceed three double-spaced pages)
   c) Listing of Present and Recent Employment, including Job Title, Location, and Dates
   d) List of Scientific Publications (Use format of J. Acoust. Soc. Am.)
   e) Honors, Awards and Society Affiliations
   f) Names and Addresses of those asked to provide Reference Letters
   g) Copy of Graduate Transcript
   h) Facilities Approval form

7. CANDIDATE'S SIGNATURE   To the best of my knowledge, the statements above and following are correct.

   _____   _____
   Signature   Date

PROPOSAL FOR RESEARCH IN ARTICULATORY VOICE ANALYSIS
AND SOURCE SEPARATION
PERRY R. COOK, CCRMA, STANFORD UNIVERSITY

This proposal concerns research on voice analysis using an articulatory model of the human vocal tract combined with other new developments in signal processing. New techniques of spectral analysis will be combined with a new model of the vocal mechanism to construct a computer system which can separate two or more speakers from a single acoustical channel, and analyze the speech sounds based on physical gestures of the vocal tract mechanism. Such a system could be applied in machine speech synthesis, automatic speech and speaker recognition, forensic acoustics, and many other areas involving speech analysis and synthesis. The proposed line of attack on the speech analysis/separation problem combines four components, each a recent and exciting topic in communications.

o   High-quality sound synthesis through physical modeling.

o   Articulatory control for accurate synthesis of speech.

o   Mapping of speech signals into an articulation-based representation.

o   Spectral decomposition of signals based on rules of grouping and cohesion.

NEW SOURCE SEGREGATION TECHNIQUES

The first phase of processing the signal will be based on recent research in polyphonic sound source segregation [Cooke91][Mellinger91][Weintraub85]. The prime cues that humans use for grouping of the spectral attributes of individual sources in a monophonic channel are common onset, common frequency modulation, common amplitude modulation, and harmonicity. In the case of voiced vowel sounds, all of these elements are present in varying degrees, allowing automatic grouping of the harmonics of the individual sources. Experiments have been performed with some success on two speakers uttering vowel passages in different pitch ranges. A technique involving an exhaustive search of time-quantized comb filter parameters to isolate the fundamental frequencies of two speakers has also shown some success [DeChevagne91]. This method has been refined in the single speaker case using techniques of adaptive comb filters [Cook91][Cook&Smith90], and could be extended to the multiple-source case.

Such techniques are successful on voiced periodic speech signals, but fail on rapidly varying source frequencies, at moments where the fundamental frequencies cross, and on unvoiced or noisy sections of speech. Speech sounds involving consonants have no analysis method proposed or investigated yet. A model of the production mechanism is needed to accurately predict and separate such acoustical events.

THE PHYSICAL VOCAL TRACT MODEL

The proposed research is centered around a dynamic vocal tract model. The physical vocal tract model has been used to produce extremely natural synthesis of the female singing and speaking voice [Cook91]. The basic attraction of such a model and its advantage over other models is that it closely represents the physical vocal tract, and if changes in the vocal tract are smooth and continuous, changes in the model are smooth and continuous. Closures in the vocal tract are a good example of a situation that is continuous in the vocal tract model but discontinuous in most other representations. The discontinuities can produce great problems in dealing with the data from spectral or other representations. Waveguide filters [Smith85] will be used to simulate the vocal production mechanism. Waveguide filters introduce no anomalies as a result of time-varying coefficients in the equations, because the filter realization computes the transmitted and reflected waves in all parts of the vocal tract. The waveguide scattering framework ensures that total system power remains constant or decreases with time.

By adapting the vocal tract dynamically to the voiced speech signal [Cook90], cues as to the production of unvoiced sounds are available. The formation of a constriction in the vocal tract model indicates the likelihood of a consonant in the future, and the place of constriction can be used to predict the consonant. Using an articulatory speech tracker, experiments have been performed which accurately track the vocal tract shape in the presence of stop consonants and fricatives. In the source segregation/analysis task, information from the extracted separated vowel trajectories would drive the articulatory vocal tract model. The articulatory information would be used to predict and identify features in the articulatory space. Synthesized spectral information at points of likely consonant production would be compared to the original analyzed spectrum, allowing the non-deterministic components to be sorted and grouped with the appropriate speaker.

OBJECTIVES

The prime objective is to use the results of this research to construct a voice source segregation and analysis system. A sub-objective is to investigate economical computational methods of performing the pitched voice source segregation task. Methods involving exhaustive searches of filter coefficient spaces, or elaborate processing of time-frequency spectrogram data might be required in the presence of more complex inharmonic sound sources, but simpler and more efficient methods might suffice for the case of voiced-speech. A second sub-objective is to refine the vocal tract adaptation algorithms. One refinement is to add physical, acoustical, linguistic, and physiological constraints on production in a given language. Other refinements

relate to economizing computation of the adaptation algorithms, using DSP chips and working toward a real-time system.

EQUIPMENT, FACILITIES, and ADVISORY PERSONNEL

CCRMA is a multidisiplinary facility with faculty, staff, and graduate students actively performing research in the areas of Musical Acoustics, Digital Signal Processing, Human and Machine Perception, Speech and Singing, Control Systems, Artificial Intelligence, Music Notation and Graphics, Composition, and other topics related to audio and music. The multidisiplinary research environment makes CCRMA an ideal facility to carry out the proposed research. A network of 18 NeXT workstations is the primary platform for teaching and research. Powerful DSP hardware and development tools are available for the NeXT machine, and these tools will be used for the proposed research. Involved with the research in an advisory capacity would be John Chowning, Professor of Music and Director of CCRMA, Julius O. Smith, Associate Professor of Music (Research), Max. V. Mathews, Professor of Music (Research), and John R. Pierce, Visiting Professor of Music Emeritus.

CONCLUSIONS

Given the current approaches to the task of multi-speaker voice source separation and analysis, only periodic sounds have been addressed, with limited success. Since the production of speech by humans is accomplished by manipulations of the human voice organ, the use of a system which tracks speech sounds in the articulatory space would yield cues as to the production of those sounds. Other approaches to computer source segregation and voice analysis suffer by ignoring this component, and a system based on an articulatory model of speech is needed to make significant progress in this area.

REFERENCES

Cook, P. R., "Speech Tracking Using an Articulatory Model of the Human Vocal Tract," CCRMA, Dept. of Music, Stanford University, May, 1990.

Cook, P. R., "Identification of Control Parameters in an Articulatory Vocal Tract Model, With Applications to the Synthesis of Singing," Elec. Engr. PhD Dissertation, Stanford, 1991.

Cook, P. R. and J.O. Smith, "An Accurate Pitch Tracker Based on Adaptive Periodic Prediction," CCRMA, Dept. of Music, Stanford University, July, 1990.

DeChevaigne', A. ., "A Mixed F0 Estimation Algorithm," ESCA, 1991.

Cooke, M. P. "Modelling Auditory Processing and Organisation'", PhD Dissertation, University of Sheffield, 1991.

Mellinger, D. K. "Event Formation and Separation in Musical Sound'", PhD Dissertation, Stanford University Computer Science, 1991.

Smith, J. O. "Waveguide Digital Filters," Center for Computer Research in Music and Acoustics (CCRMA), Dept. of Music, Stanford University, March 1985.

Weintraub, M. "A theory and computational model of auditory monaural sound separation," PhD. Dissertation, Stanford University Electrical Engineering, 1985.