

Section 2 Sensory Communication

Chapter 1 Sensory Communication

Chapter 1. Sensory Communication

Academic and Research Staff

Professor Louis D. Braid, Nathaniel I. Durlach, Professor Richard M. Held, Dr. Paul Duchnowski, Dr. Julie E. Greenberg, Dr. Matthew H. Power, Dr. William M. Rabinowitz, Dr. Charlotte M. Reed, Dr. Kourosh Saberi, Dr. J. Kenneth Salisbury, Dr. Barbara G. Shinn-Cunningham, Dr. Mandayam A. Srinivasan, Dr. Hong Z. Tan, Dr. Thomas E.v. Wiegand, Dr. David Zeltzer, Dr. Patrick M. Zurek, Merry A. Brantley, Andrew R. Brughera, Lorraine A. Delhorne, Ann K. Dix, Dorrie Hall, Seth M. Hall, David S. Lum, Coral D. Tassa

Visiting Scientists and Research Affiliates

Dr. Cagatay Basdogan, Dr. G. Lee Beauregard, Dr. Susan L. Goldman, Dr. Kenneth W. Grant, Dr. Karen L. Payton, Geoffrey L. Plant, Dr. Christine M. Rankovic, Dr. Kaoru Sekiyama, Dr. Anne H. Takeuchi

Graduate Students

Walter A. Aviles, Douglas S. Brungart, Jyh-Shing Chen, Suvranu De, Joseph G. Desloge, Joseph A. Frisbie, Isaac Graaf, Chih-Hao Ho, I-Chun A. Hou, Steingrimur P. Karason, Rabi Karmacharya, Salim F. Kassem, Glenn Koh, Jean C. Krause, Jung-Chi Liao, Gregory G. Lin, Sharieff A. Mansour, Kinuko Masaki, Amanda S. Myers, Michael P. O'Connell, John Park, Balasundar I. Raju, David W. Schloerb, Steven J. Schlueter, Matthew G. Sexton, Jason Sroka, Prasanna B. Tambe, Kimberly J. Voss, Elias A. Vyzas, Evan F. Wies, Elron A. Yellin

Undergraduate Students

Dylan J. Birtolo, David E. DiFranco, Alexis E. Farel, Matthew B. Keller, Bryan D. Kincy, Kari Anne H. Kjolaas, Lajos Molnar, Sadiki P. Mwanyoha, T.H. Ogora, Jonathan D. Pfautz, Deborah P. Rhoads, Jonathan R. Santos, Ranjini Srikantiah, Adrienne H. Slaughter, Maciej Stachowiak

Technical and Support Staff

Rebecca L. Garnett, Eleanora M. Luongo, Denise M. Rossetti, Francis G. Taylor

1.1 Introduction

The Sensory Communication group is interested in understanding sensorimotor and cognitive processes and exploiting this understanding to solve practical problems in a variety of application domains. Our basic research is characterized by behavioral (psychophysical) experimentation and quantitative theoretical modeling. Facilities to support this research are developed as needed. Although some research is conducted on vision, most of the group's work is focused on audition or tacton. The main applications are concerned with aiding individuals who suffer from impaired hearing, developing improved human-machine interfaces for virtual environments and teleoperation (virtual reality), and the use of virtual environment technology for training. The main facilities of the group are associated with the psychoacoustics, touch (or haptics), and virtual environment laboratories.

In the following report, research is organized according to the sources of funding provided to our group. Thus, work on hearing and touch, and to a certain extent vision, is described in a variety of subsections in a variety of funding contexts.

1.2 Hearing Aid Research

Sponsor

National Institutes of Health
Grant RO1 DC00117

Project Staff

Professor Louis D. Braid, Dr. Paul Duchnowski, Dr. Karen L. Payton, Dr. Matthew H. Power, Dr. Christine M. Rankovic, Dr. Charlotte M. Reed, Dr. Patrick M. Zurek, Isaac Graaf, Jean C. Krause, David S. Lum, Sadiki P. Mwanyoha, Steven J. Schlueter, Jason Sroka

1.2.1 Specific Aims

Our long-term goal is to develop improved hearing aids for people suffering from sensorineural hearing impairments. Our efforts are focused on problems resulting from inadequate knowledge of the effects of various alterations of speech signals on speech reception by impaired listeners, specifically on the fundamental limitations on the improvements in speech reception that can be achieved by processing speech. Our aims are:

1. To assess the relative contributions of various functional characteristics of hearing impairments to reduced speech-reception capacity.
2. To evaluate the effects of style of speech articulation and variability in speech production on speech reception by hearing impaired listeners.
3. To develop and evaluate analytical models that can predict the effects of a variety of alterations of the speech signal on intelligibility.
4. To develop and evaluate signal processing techniques that hold promise for increasing the effectiveness of hearing aids.

1.2.2 Studies and Results

Characteristics of Sensorineural Hearing Impairment

Simulation of Sensorineural Hearing Loss

Although both the design and selection of hearing aids are informed by the results of research on auditory physiology and perception, the processes are typically hampered by the problem of obtaining insight into the effects of various signal processing transformations on the auditory stimulus presented to the impaired listener. To help overcome this problem we have been developing *functional*

models that allow listeners with normal hearing to experience the perceptual distortions and speech-reception problems experienced by hearing-impaired listeners. Past work¹ has focussed on simulations based on masking noise and multiband amplitude expansion.² Both of these simulations are capable of elevating the tone-detection thresholds of normal-hearing listeners to those of listeners with hearing impairments, as well as inducing the abnormal growth of loudness experienced by listeners with sensorineural impairments. Both of these approaches also simulate the effects of mild and moderate hearing impairments on speech reception. The effects of more severe impairments are not always well simulated. In particular, listeners with simulated impairments perform roughly as well in tests of speech reception in noise as listeners with severe hearing impairments when amplification with a flat frequency-gain characteristic is used, but not when high-frequency emphasis is provided.

One factor that may be responsible for this discrepancy is reduced frequency selectivity, i.e., the increased sensitivity of impaired listeners to out-of-band noise. Both simulations based on masking noise and those based on amplitude expansion have been shown to reproduce some aspects of reduced frequency selectivity, but not others.³ In particular, both simulations result in widened psychophysical tuning curves, as measured in a forward masking paradigm, but neither broadens the pattern of simultaneous masking produced by an intense narrowband noise.

In current research, we are evaluating the effect of combining hearing loss simulations that elevate threshold and induce abnormal loudness growth with simulations that reduce frequency selectivity by smearing the short-term spectrum. The goal of these combinations is to produce roughly the same excitation patterns in a normal ear as the unprocessed stimuli would evoke in an impaired ear. Elevated thresholds and recruitment were simulated

¹ P.M. Zurek and L.A. Delhorne, "Consonant Reception in Noise by Listeners with Mild and Moderate Hearing Impairment," *J. Acoust. Soc. Am.* 82: 1548-1559 (1987); P. Duchnowski and P.M. Zurek, "Villchur Revisited: Another Look at AGC Simulation of Recruiting Hearing Loss," *J. Acoust. Soc. Am.* 98: 3170-3181 (1995); S.V. DeGennaro and L.D. Braida, "Effects of Linear Amplification and Amplitude Compression on Sentence Reception by Listeners with Hearing Loss Simulated by Masking Noise," in *Modelling Sensorineural Hearing Loss*, ed. W. Jesteadt (Mahwah, New Jersey: L. Earlbaum Assoc., 1997); D.S. Lum and L.D. Braida, "DSP Implementation of a Real-Time Hearing Loss Simulator Based on Dynamic Expansion," in *Modelling Sensorineural Hearing Loss*, ed. W. Jesteadt (Mahwah, New Jersey: L. Earlbaum Assoc., 1997).

² E. Villchur, "Electronic Models to Simulate the Effect of Sensory Distortions on Speech Perceptions by the Deaf," *J. Acoust. Soc. Am.* 62: 665-674 (1977).

³ D.S. Lum, *Evaluation of a Hearing-Loss Simulator*, S.B. thesis, Dept. of Electr. Eng. and Comput. Sci., MIT, 1994.

via multiband amplitude expansion of the signal.⁴ Reduced frequency selectivity was simulated by smearing the power spectrum of the signal.

The effects of the simulations were evaluated in psychoacoustic experiments that measured simultaneous masking patterns produced by narrowband noise, forward masked psychophysical tuning curves, and loudness summation for bands of noise.

The results of the two masking experiments showed that in isolation both the expansion simulation and spectral smearing simulation were capable of reducing frequency selectivity substantially. However, in combination, the simulations produced a more accurate simulation of the excessive spread of masking exhibited by some hearing impaired listeners. Each simulation produced wider forward masked PTCs, but the spectral smearing simulation produced considerably more broadening in the high frequency portion of the tuning curve than did the expansion simulation. In both masking experiments, combining the two simulations produced effects that were smaller than additive. The expansion simulation tended to reduce loudness summation, but spectral smearing had little effect on loudness summation.

These results suggest that both recruitment and poor frequency selectivity adversely affect auditory perception in listeners with hearing impairments. Moreover, compensating for one degradation may exacerbate the effects of the other.

Characteristics of the Speech Signal

Clear Speech

In adverse listening conditions, talkers can increase their intelligibility by speaking clearly. While producing clear speech, however, talkers often significantly reduce their speaking rate. A recent study⁵ showed that talkers can be trained to produce a form of clear speech at normal rates (clear/normal speech). This finding suggests that acoustical factors other than reduced speaking rate are responsible for the high intelligibility of clear speech. To gain insight into these factors, the acoustical properties of conversational, clear/slow,

and clear/normal speech are being analyzed to determine phonological and phonetic differences between the two speaking modes.

As in an earlier study of the acoustics of clear speech,⁶ acoustical measurements for two talkers were made at three levels of detail: global, phonological, and phonetic. At the global level, differences in pause duration and long-term RMS spectra were examined. At the phonological level, the frequency of various phonological phenomena that occur in running speech were measured. These phenomena are classified into the following categories: vowel modification, burst elimination, alveolar flap, and sound insertion. Phonetic level measurements included segmental phone duration, segmental power, and short-term RMS spectra. Initial acoustical measurements of conversational and clear/slow speech are consistent with Picheny et al.'s data. Measurements for clear/normal speech indicate that talkers have different strategies for producing highly intelligible speech at normal rates. Each talker has retained a different subset of the acoustical characteristics that are typical of clear/slow speech.

In order to evaluate the robustness of the different strategies for producing clear/normal speech, additional intelligibility tests have been used to investigate other degraded environments. Previous measurements have shown that both talkers can produce clear/normal speech which is highly intelligible in additive noise for normal hearing listeners. Four additional environments have now been explored: normal hearing listeners were tested in reverberation, low-pass filtering, and high-pass filtering conditions, and non-native listeners were tested in additive noise. Preliminary results showed a significant talker effect for intelligibility scores in other environments. This result suggests some talkers may have more robust strategies for producing clear/normal speech in a variety of degraded environments.

We plan to characterize the acoustical differences between clear and conversational speech produced at the same speaking rate. Our new database of clear and conversational speech will be phonetically segmented and labelled. Several short-term acoustic differences between clear and conversational

⁴ D.S. Lum and L.D. Braid, "DSP Implementation of a Real-Time Hearing Loss Simulator Based on Dynamic Expansion," in *Modelling Sensorineural Hearing Loss*, ed. W. Jesteadt (Mahwah, New Jersey: L. Earlbaum Assoc., 1997).

⁵ J.C. Krause, *The Effects of Speaking Rate and Speaking Mode on Intelligibility*, S.M. thesis, Dept. Electr. Eng. and Comput. Sci., MIT, 1995.

⁶ M.A. Picheny, N.I. Durlach, and L.D. Braid, "Speaking Clearly for the Hard of Hearing II: Acoustic Characteristics of Clear and Conversational Speech," *J. Speech Hear. Res.* 29: 434-446 (1986).

speech will be investigated for their effect on intelligibility. For example, the consonant-to-vowel ratios (CVR) found in clear speech are substantially greater than those in conversational speech.⁷ Recently, two studies have reported an intelligibility improvement due to artificial CVR increase for nonsense syllables presented to impaired listeners⁸

Modeling the Effect of Speaking Style Intelligibility

Although currently accepted models of speech intelligibility account for the effects of linear filtering and additive noise on intelligibility fairly well, at least for listeners with normal hearing, they cannot account for a number of signal transformations of importance to listeners with hearing impairments. These more problematic transformations are not easily described in terms of changes in the long-term spectrum of speech or changes in the signal to noise ratio. For example, whereas clear and conversational speech have roughly the same long-term spectrum,⁹ they are likely to have different modulation spectra because of differences in speaking rate and consonant-vowel energy (CV) ratios. Similarly, the effects of reverberation and multiband amplitude compression on the speech signal are likely to be more evident in the intensity modulation spectra than in the power spectrum.

Current work on this project is directed at using the speech transmission index (STI)¹⁰ to understand the dependence of intelligibility on speaking. At present the STI is measured or computed using a representation of the modulation patterns of conversational speech. Different speaking styles require that different modulation patterns be specified. We are developing a method of calculating the speech transmission index from measurements made on real speech waveforms. However these measure-

ments have proved difficult to make because of noise-induced artifacts in speech envelope spectra. Although we have attempted to develop techniques to identify these artifacts so that they can be removed from the measurements, our efforts have met with only limited success. We are currently studying means to predict when the artifacts would dominate the high-frequency envelope spectra and use the predictions to specify an upper modulation frequency limit for the measurements. We found that reliable modulation transfer functions (MTFs) can be computed from speech envelope spectra if we computed the coherence function and used it to indicate a modulation frequency limit.¹¹ A coherence criterion level of 0.8 was found to be appropriate for clear and conversational speech in noise. This same criterion level was also found to truncate reverberant speech envelope spectra in both speaking styles at appropriate modulation frequencies to remove the artifacts seen there.

We have computed speech transmission indices (STIs) from the modified modulation transfer functions (MTFs) that were derived by Payton and Braida.¹² Initial comparisons suggest, for the original four conditions evaluated, there were distinct and nonuniform differences between STI computations based on theoretical calculations and computations that are based on the envelope spectra of real speech. In particular, the lowest speech octave bands showed enhancement of some low modulation frequencies in the presence of reverberation. In an attempt to understand this phenomenon better, MTFs and STIs for all 9 environmental conditions are being computed for each speaking style. This computationally intensive effort is almost completed. The correlation of the new complete set of data with the experimental results will be compared with the correlation observed using the data of Payton et al.¹³ When this analysis is complete, we

⁷ M.A. Picheny, 1981.

⁸ A.A. Montgomery and R.A. Edge, "Evaluation of Two Speech Enhancement Techniques to Improve Intelligibility for Hearing-impaired Adults," *J. Speech Hear. Res.* 29: 434-446 (1986); S. Gordon-Salant, "Effects of Acoustic Modification on Consonant Perception by Elderly Hearing-impaired Subjects," *J. Acoust. Soc. Am.* 77: S106 (1986).

⁹ M.A. Picheny, N.I. Durlach, and L.D. Braida, "Speaking Clearly for the Hard of Hearing II: Acoustic Characteristics of Clear and Conversational Speech," *J. Speech Hear. Res.* 29: 434-446 (1986).

¹⁰ T. Houtgast and H.J.M. Steeneken, "A Review of the MTF Concept in Room Acoustics and Its Use for Estimating Speech Intelligibility in Auditoria," *J. Acoust. Soc. Am.* 77: 1069-1077 (1985).

¹¹ K.L. Payton and L.D. Braida, "Speech Modulation Transfer Functions for Different Speaking Styles," *J. Acoust. Soc. Am.* 98: 2982 (1995).

¹² K.L. Payton and L.D. Braida, "Speech Modulation Transfer Functions for Different Speaking Styles," *J. Acoust. Soc. Am.* 98: 2982 (1995).

¹³ K.L. Payton, R.M. Uchanski, and L.D. Braida, "Intelligibility of Conversational and Clear Speech in Noise and Reverberation for Listeners with Normal and Impaired Hearing," *J. Acoust. Soc. Am.* 95: 1581-1592 (1994).

plan to investigate the envelope spectra, MTFs and STIs of amplitude compressed speech and amplitude compressed speech in noise.

Computational Models of Intelligibility

Power and Braid¹⁴ developed a computational model of speech intelligibility and tested its ability to predict the effects of filtering and additive noise on speech perception. In this model intelligibility is assumed to be monotonically related to the mutual information between a symbol sequence corresponding to a perceptual representation of undegraded speech waveforms and a symbol sequence representing the degraded waveform. This model provides a relatively good account of the effect of lowpass and highpass filtering on intelligibility. However, the predicted effect of additive noise is much greater than is typically observed for human listeners.

This excessive sensitivity of the model to the effects of additive noise is similar to that seen in the performance of Automated Speech Recognition (ASR) systems in degraded signal conditions. It has been suggested that if the physiological processing stages of human processing were represented in automated approaches, ASR systems would exhibit levels of performance and robustness similar to those achieved by humans. It seems likely that the same strategy would improve our ability to predict speech intelligibility for human listeners.

We are attempting to determine the effect of incorporating functional models of human processing stages into ASR systems. Whereas most speech recognition studies only examine final recognition rates, this study will examine the error patterns of traditional ASR approaches as well as systems that incorporate functional models of human auditory processing. The error patterns will be compared with human error patterns to determine if incorporation of the functional models makes ASR systems perform more similarly to humans. This approach should increase our understanding of the contribution those stages make in the human speech recognition process.

1.2.3 Publications

DeGennaro, S.V., and L.D. Braid. "Effects of Linear Amplification and Amplitude Compression on Sentence Reception by Listeners with Hearing Loss Simulated by Masking Noise." In *Modelling Sensorineural Hearing Loss*. Ed. W. Jesteadt. Mahwah, New Jersey: L. Earlbaum Assoc., 1997.

Lum, D.S., and L.D. Braid. "DSP Implementation of a Real-Time Hearing Loss Simulator Based on Dynamic Expansion." In *Modelling Sensorineural Hearing Loss*. Ed. W. Jesteadt. Mahwah, New Jersey: L. Earlbaum Assoc., 1997.

Power, M.H., and L.D. Braid. "Consistency among Speech Parameter Vectors: Application to Predicting Speech Intelligibility." *J. Acoust. Soc. Am.* 100: 3882-3898 (1996).

Uchanski, R.M., S. Choi, L.D. Braid, C.M. Reed, and N.I. Durlach. "Speaking Clearly for the Hard of Hearing IV: Further Studies of the Role of Speaking Rate." *J. Speech Hearing Res.* 39: 494-509 (1996).

1.3 Enhanced Communication for Speechreaders

Sponsor

National Institutes of Health
Grant RO1 DC02032

Project Staff

Professor Louis D. Braid, Dr. Paul Duchnowski, Dr. Susan L. Goldman, Dr. Matthew H. Power, Dr. Anne H. Takeuchi, Lorraine A. Delhorne, Dr. Charlotte M. Reed, Dr. Kaoru Sekiyama, Dr. Kenneth W. Grant, Frederick W. Chen, Joseph A. Frisbie, Matthew B. Keller, David S. Lum, Sharieff A. Mansour, Deborah P. Rhoads, Matthew G. Sexton, Prasanna B. Tambe

¹⁴ M.H. Power and L.D. Braid, "Consistency among Speech Parameter Vectors: Application to Predicting Speech Intelligibility," *J. Acoust. Soc. Am.* 100: 388-3898 (1996).

1.3.1 Specific Aims

Although speechreading is an essential component of communication for the hearing impaired under nearly all conditions, the ability to communicate through speechreading alone is severely constrained because many acoustic distinctions important to speech reception are not manifest visually. Supplements derived from the acoustic speech signal have been shown to improve speech reception markedly when the cues they provide are integrated with cues derived from the visible actions of the talker's face. Our long-term goal is to develop aids for individuals with hearing impairments so severe that their communication relies heavily on speechreading. Our specific aims are:

1. To develop models of audiovisual integration to quantify how well supplementary signals are integrated with speechreading.
2. To develop and evaluate simplified acoustic signals that can be derived from acoustic speech by signal processing. Such signals would form the basis of new types of hearing aids for listeners with severe hearing impairments.
3. To develop systems for producing and displaying discrete speechreading supplements similar to the "manual cued speech" system that can be derived from the acoustic signal by speech recognition technology. Such supplements would display streams of discrete symbols that would be derived automatically from acoustic speech and presented visually for integration with the speechreading signal.

1.3.2 Studies and Results

Supplements Based on Signal Processing

Although many signals that supplement speechreading effectively can be derived from the acoustic speech signal, supplements based on the amplitude envelopes of filtered bands of speech are easily extracted, relatively resistant to background noise, and readily integrated with speechreading, at least by listeners with normal-hearing.

In past work, Nadeau¹⁵ programmed wearable DSP-based SiVo Aids¹⁶ to produce speechreading supplements consisting of one or more tones amplitude modulated by the amplitude envelopes of bands of speech. Of two listeners with severe to profound hearing impairments who were fitted with the aids, one derived very little benefit and discontinued using the aid after ten weeks of use while the other derived modest benefits and continues to use the aid after 18 months of use. During the past year, we screened an additional ten impaired listeners, but were able to identify only one who seemed likely to benefit from the envelope signals. After being fitted with a wearable aid that produced these signals, however, this listener was unwilling to use the aid in the field for more than one month.

The limited acceptance of this type of speechreading aid by the hearing impaired most likely stems from two causes: cosmetic factors and the limited benefit provided by the envelope signals themselves. In future work, we plan to evaluate other wearable devices capable of providing the envelope-based signals. The benefit provided by the envelope-based signals most likely stems from the difficulty of providing more than one such signal in the limited auditory area of the impaired listener. Grant et al.¹⁷ have shown that, while two such signals can provide listeners with normal hearing with considerably more benefit than a single envelope-based signal, the increase depends upon presenting the modulated tones at frequencies that would be inaudible to most listeners with severe hearing losses. Since listeners with hearing impairments typically have little residual hearing at high frequencies, they are unable to benefit from signals that convey the envelopes of high frequency bands

¹⁵ P.M. Nadeau, *Amplitude Envelope Cues as an Aid to the Speechreading of Impaired Individuals*, S.M. thesis, Dept. of Electr. Eng. and Comput. Sci., MIT, 1995.

¹⁶ A. Faulkner, V. Ball, S. Rosen, B.C.J. Moore, and A. Fourcin, "Speech Pattern Hearing Aids for the Profoundly Hearing Impaired: Speech Perception and Auditory Abilities," *J. Acoust. Soc. Am.* 91: 2136-2155 (1992).

¹⁷ K.W. Grant, L.D. Braida, and R.J. Renn, "Auditory Supplements to Speechreading: Combining Amplitude Envelope Cues from Different Spectral Regions of Speech," *J. Acoust. Soc. Am.* 95: 1065-1073 (1994).

of speech. To overcome this problem, we are developing and evaluating techniques for conveying the cues provided by the envelopes of high frequency bands to listeners with low frequency hearing.

Supplements Based on Automatic Speech Recognition

In manual cued speech (MCS) a speaker gestures with his/her hand to resolve ambiguities among speech elements that are often confused by speechreaders. The shape of the hand distinguishes among consonants, and the position of the hand relative to the face distinguishes among vowels. Experienced receivers of MCS achieve nearly perfect reception of everyday connected speech.

Visual supplements related to the manual cued speech system used in the education of the deaf have been shown to improve speechreading dramatically.¹⁸ These supplements can be derived using the techniques of automatic speech recognition (ASR), a technology that has advanced rapidly during the past decade. Automatically generated cues would not require a speaker to be skilled in cue production, and so could assist a wider variety of individuals in everyday situations. Two major hurdles must be overcome to develop an effective automatic cueing system: the ASR technology used must provide cues with sufficient accuracy; and the cues must be displayed in a manner that can be integrated well with speechreading.

Based on the results of our simulation studies,¹⁹ we have begun to develop and evaluate real-time cueing systems based on automatic speech recognizers. The cues presented by these systems attempt to reproduce the coding rules familiar to users of manual cued speech faithfully so that experienced cue receivers can evaluate their effectiveness. The initial system uses a speaker-dependent phonetic recognizer based on the Entropic HTK software package. A real-time acoustic front end that derives mel-frequency cepstral coefficients at a frame rate of 100/sec. has been implemented using DSP hardware. Cues are displayed on a video monitor as images of hand

shapes located at appropriate positions around the face of the speaker.²⁰ At present these operations are carried out by three laboratory computers. Special efforts have been made to ensure that the display of cues is properly synchronized with the facial actions of the talker.

Initial studies indicate that only minimal aid is provided to the speechreader when cues are derived by an HMM recognizer based on context-independent phone models. However, consistent benefits are observed when recognition is based on context-dependent phone models. We are currently unable to implement such a recognizer in real time using existing laboratory computers. Future research is aimed at developing an automatic speech recognition system with the speed, accuracy, and computational efficiency required for a real-time automatic cueing system.

1.3.3 Publication

Bratakos, M.S., P. Duchnowski, and L.D. Braida. "Toward the Automatic Generation of Cued Speech." Submitted to *Cued Speech J.*

Thesis

Sexton, M. *A Video Display System for an Automatic Cue Generator*. M.Eng. thesis, Dept. of Electr. Eng. and Comput. Sci., MIT, 1997.

1.4 Tactile Communication of Speech

Sponsor

National Institutes of Health/National Institute on Deafness and Other Communication Disorders
Grant R01 DC00126

Project Staff

Lorraine A. Delhorne, Nathaniel I. Durlach, Alexis E. Farel, Seth M. Hall, Chih-Hao Ho, Geoffrey L. Plant, Dr. William M. Rabinowitz, Dr. Charlotte M. Reed, Dr. Mandayam A. Srinivasan, Dr. Hong Z. Tan, Jonathan R. Santos

¹⁸ R.M. Uchanski, L.A. Delhorne, A.K. Dix, L.D. Braida, C.M. Reed, and N.I. Durlach, "Automatic Speech Recognition to Aid the Hearing Impaired: Prospects for the Automatic Generation of Cued Speech," *J. Rehab. Res. Dev.* 31: 20-41 (1994).

¹⁹ M.S. Bratakos, P. Duchnowski, and L.D. Braida, "Toward the Automatic Generation of Cued Speech," submitted to *Cued Speech J.*

²⁰ M. Sexton, *A Video Display System for an Automatic Cue Generator*, M. Eng. thesis, Dept. of Electr. Eng. and Comput. Sci., MIT, 1997.

1.4.1 Goals and Significance

This research is directed towards the development of effective tactual communication devices for individuals with profound deafness or deaf-blindness. Such devices would lead to improved speech reception, speech production, language competence, and awareness of environmental sounds for such individuals and would provide them with a sensory-aid option in addition to hearing aids and cochlear implants. At a more basic scientific level, this research contributes to increased understanding of speech communication, environmental-sound reception, tactual perception, manual sensing, display design, and sensory substitution.

1.4.2 Studies and Results over the Past Year

Basic Studies of Hand Stimulation and Active Touch

Progress during the current year includes the completion of a doctoral thesis concerned with the transfer of information through the tactual sense using a newly developed multifinger positional display²¹ and the completion of a master's thesis on the discrimination of thickness.²² Manuscripts describing the research conducted in both theses are currently being prepared for publication. In addition, studies on additivity of information and forward/backward masking effects have been conducted with our new multifinger positional display.²³ A preliminary test of the additivity theory proposed by Durlach et al.²⁴ for multidimensional information transfer (IT) was conducted using the dimensions of frequency and amplitude presented to the index finger. For experiments in which the subject's task

was to identify either frequency alone (from a set of eight values in the region of 100-300 Hz) or amplitude alone (from a set of eight values in the range of 12-40 dB SL), estimates of IT obtained with roving values of the non-relevant background dimension were lower than those obtained when those values were fixed. The ratio of roving-to-fixed IT measurements was 0.51 for frequency versus 0.70 for amplitude identification. Estimates of IT were also obtained in an identification experiment where the subject's task was to identify both the frequency and amplitude of 64 signals (formed from the combination of eight frequencies with eight amplitudes). The estimate of IT for the two-dimensional identification task (corrected for response bias arising from the number of trials used to calculate IT), 1.29 bits, was quite similar to the sum of the IT estimates for frequency and amplitude identification using roving background parameters (0.38 bits and 0.88 bits, respectively), as predicted by the theory of Durlach et al.²⁵ Finally, a study has been conducted to extend Tan's²⁶ research on the identification of multidimensional signals in the presence of backward and forward maskers. While Tan's masking paradigm required identification of a target signal in the presence of both a backward and a forward masker, the present study examined the separate effects of each type of masker on the identification of a set of 30 250-msec multidimensional signals presented to the index finger. The masker, which also had a fixed duration of 250 msec, was presented at five different inter-stimulus intervals ranging from 25-400 msec for either forward (masking preceding target) or backward (target preceding masker) conditions. The results were qualitatively similar to those obtained in previous tactile recognition masking studies conducted by Craig and Evans²⁷ using very brief line patterns presented to the index finger. That is,

²¹ H.Z. Tan, *Information Transmission with a Multi-Finger Tactual Display*, Ph.D. diss., Dept. of Electr. Eng. and Comput. Sci., MIT, 1996.

²² C-H. Ho, *Human Haptic Discrimination of Thickness*, S.M. thesis, Dept. Mech. Eng., MIT, 1996.

²³ H.Z. Tan and W.M. Rabinowitz, "A New Multi-Finger Tactual Display," *Proceedings of the American Society of Mechanical Engineering (ASME) Dynamics Systems and Control Division*, ASME 1996, DSC-Vol. 58: 515-522.

²⁴ N.I. Durlach, H.Z. Tan, N.A. Macmillan, and W.M. Rabinowitz, "Resolution in One Dimension with Random Variations in Background Dimensions," *Percept. Psychophys.* 46: 293-296 (1986).

²⁵ N.I. Durlach, H.Z. Tan, N.A. Macmillan, and W.M. Rabinowitz, "Resolution in One Dimension with Random Variations in Background Dimensions," *Percept. Psychophys.* 46: 293-296 (1986).

²⁶ H.Z. Tan, *Information Transmission with a Multi-Finger Tactual Display*, Ph.D. diss., Dept. of Electr. Eng. and Comput. Sci., MIT, 1996.

²⁷ J.C. Craig and P.M. Evans, "Temporal Integration and Vibrotactile Backward Masking," *J. Exper. Psychol.: Human Percept. Perform.* 12: 160-168 (1986); J.C. Craig and P.M. Evans, "Vibrotactile Masking and the Persistence of Tactual Features," *Percept. Psychophys.* 42: 309-317 (1987).

backward masking effects were greater than forward masking effects at short values of stimulus-onset asynchrony; however, the effects of forward masking appeared to persist for a longer period of time. These results are being compared to those obtained by Tan and will play a role in developing a more complete understanding of the processes involved in recognition of tactual stimulus streams.

Evaluation of Wearable Tactile Aids

Research conducted during the current year includes studies of speech reception and/or speech production in both children and adults fitted with wearable tactile aids. A program is under development for teaching speech to children ranging in age from three to ten years and who are attending schools where American Sign Language or Signed English is the primary mode of communication. Work has also been conducted on the development of a simple screening protocol for children in the earliest stages of speech development. A study of the speech intelligibility of a congenitally deaf young adult who was provided with speech training following the fitting of a tactile aid is nearly completed. Tests of the intelligibility of consonants in nonsense syllables and sentences produced by this subject indicated that normal-hearing listeners had higher scores on post-training productions compared to pre-training productions. Pre- and post-training scores were 26.2 percent compared to 54.2 percent for consonants, respectively, and 8.4 percent compared to 29.5 percent for sentences, respectively. In addition, a script has been developed for a set of video-taped training sessions for adults fitted with tactile aids, providing practice in recognition of target consonants and vowels in meaningful words and sentences as well as synthetic exercises.

Development of Improved Tactual Supplements to Speechreading

Work in this area has been concerned with evaluation of the reception of voicing through various tactile displays. A detailed examination of the discriminability of seven voiced-voiceless contrasts was conducted on a highly experienced user of the Tactaid 7 device. Differences in performance were observed across individual contrasts and will be examined to determine which cues to voicing appear to be most salient. In addition, work is being conducted to design a signal-processing scheme for presenting speech cues through our new multifinger positional display which will take advantage of its multidimensional capabilities. Finally, we have received delivery of two "Tickle Talker" devices from the University of Melbourne

and have begun to measure thresholds and discomfort levels on members of our research staff for the electrotactile stimulation provided by this device.

Study of the Reception of Environmental Sounds through Tactual Displays

Software has been developed for conducting experiments on the identification of environmental sounds. For each of four contexts, a set of ten appropriate sounds has been selected from signals available on CD-ROM recordings. These signals have been edited to provide three separate tokens of each sound with a duration that is "natural" for a given type of environmental sound (across sounds, durations range from roughly 250 msec for a slamming door to four seconds for an appliance motor). The software controls signal presentation and stores subject responses (using a mouse to click on a response selected from a closed-set of alternatives).

1.4.3 Publications

- Delhorne, L.A., J.M. Besing, N.I. Durlach, and C.M. Reed. "Tactual Cued Speech as a Supplement to Speechreading." *Cued Speech J.* Forthcoming.
- Plant, G. "Tactrain: Tactaid VII Training Program for Adults." Somerville, Massachusetts: Hearing Rehabilitation Foundation, 1996.
- Plant, G., M. Horan, and H. Reed. "Speech Teaching for Deaf Children in the Age of Bilingual/Bicultural Programs: The Role of Tactile Aids." *Proceedings of ISAC-96 (International Sensory Aid Conference)*. Eds. N. Van Son and F. Coninx. Instituut voor Doven, Sint-Michielsgestel, The Netherlands, 1996. (Also invited for submission in special issue of *Scandinavian Audiol.*)
- Reed, C.M. "Implications of the Tadoma Method of Speechreading for Spoken Language Processing." *Proceedings of the Fourth International Conference on Spoken Language Processing (ICSLP96)*, Volume 3: SaA1S1-T4, 1996.
- Reed, C.M., and N.I. Durlach, N.I. "Note on Information Transfer Rates in Human Communication." Submitted to *Presence*.
- Tan, H.Z., and W.M. Rabinowitz. "A New Multifinger Tactual Display." *Proceedings of the American Society of Mechanical Engineers*

Dynamics Systems and Control Division, ASME
1996, DSC-Vol. 58: 515-522.

Tan, H.Z., N.I. Durlach, W.M. Rabinowitz, and C.M. Reed. "Information Transmission with a Multi-Finger Tactual Display." *Proceedings of ISAC-96 (International Sensory Aid Conference)*. Eds. N. Van Son and F. Coninx. Instituut voor Doven, Sint-Michielsgestel, The Netherlands. (Also invited for submission in special issue of *Scandinavian Audiol.*)

Tan, H.Z., N.I. Durlach, W.M. Rabinowitz, C.M. Reed, and J.R. Santos. "Reception of Morse Code through Motional, Vibrotactile, and Auditory Stimulation." *Percept. Psychophys.* Forthcoming.

Theses

Farel, A.E. *Recognition Masking Paradigm*. Advanced Undergraduate Paper. Dept. of Electr. Eng. and Comput. Sci., MIT, 1996.

Ho, C-H. *Human Haptic Discrimination of Thickness*. S.M. thesis, Dept. of Mech. Eng., MIT, 1996.

Tan, H.Z. *Information Transmission with a Multi-Finger Tactual Display*. Ph.D. diss., Dept. of Electr. Eng. and Comput. Sci., MIT, 1996.

1.5 Multimicrophone Hearing Aids

Sponsor

National Institutes of Health
Grant R01 DC00270

Project Staff

Andrew R. Brughera, Joseph G. Desloge, Dr. Julie E. Greenberg, Michael P. O'Connell, Dr. William M. Rabinowitz, Coral D. Tassa, Dr. Patrick M. Zurek

1.5.1 Overview

The goal of this research is to determine the improvements that can be provided to hearing aids through the use of multiple microphones. The work is directed toward developing algorithms for processing the signals from a head-worn microphone array for the primary goal of improving the intelligibility of speech (assumed to arise from a known direction) in the presence of noise and reverberation. Ideally, this intelligibility enhancement would be achieved without compromising the listener's ability to monitor and localize sound sources from all directions. Array processing algorithms are first implemented and evaluated in terms of signal-to-noise improvement in computer simulations. The most promising approaches are then implemented in real-time with wearable devices (tethered to a computer) for laboratory evaluations in terms of speech reception in noise and sound localization by normal and hearing-impaired listeners.

Recent work has focused on developing microphone array hearing aids with binaural outputs that allow users to localize sound sources.²⁸ Current projects include (1) a new very-rapidly-adapting null-steering algorithm; and (2) assessment of sound fields in everyday environments to determine optimal processing schemes and eventual performance.

1.5.2 Publications

Desloge, J.G., W.M. Rabinowitz, and P.M. Zurek. "Microphone-array Hearing Aids with Binaural Output. I. Fixed-processing Systems." *IEEE Trans. Speech Audio Proc.* Forthcoming.

Welker, D.P., J.E. Greenberg, J.G. Desloge, and P.M. Zurek. "Microphone-array Hearing Aids with Binaural Output. II. A Two-microphone Adaptive System." *IEEE Trans. Speech Audio Proc.* Forthcoming.

²⁸ J.G. Desloge, W.M. Rabinowitz, and P.M. Zurek, "Microphone-array Hearing Aids with Binaural Output. I. Fixed-processing Systems," *IEEE Trans. Speech Audio Proc.*, forthcoming; D.P. Welker, J.E. Greenberg, J.G. Desloge, and P.M. Zurek, "Microphone-array Hearing Aids with Binaural Output. II. A Two-microphone Adaptive System," *IEEE Trans. Speech Audio Proc.*, forthcoming.

1.6 Hearing Aid Device Development

Sponsor

National Institutes of Health
Contract N01 DC52107

Project Staff

Merry A. Brantley, Andrew R. Brughera, Dr. Julie E. Greenberg, Dr. William M. Rabinowitz, Coral D. Tassa, Elron A. Yellin, Dr. Patrick M. Zurek

1.6.1 Overview

The overall objective of work under this contract is to evaluate promising signal processing algorithms for hearing aids under realistic conditions. Progress toward this goal has begun with development of a laboratory-based signal-processing and testing system. Subsequent laboratory evaluations will select signal processing algorithms for implementation in wearable devices, which are being designed and built by Sensimetrics Corporation. Later in the project, field studies of hearing-impaired persons using this device will be conducted to evaluate the effectiveness of algorithms aimed at improving speech reception in background noise, preventing loudness discomfort, and increasing maximum gain without feedback.

Recent work has been focused on test development and on defining fitting strategies for linear gain response. Approximately twelve users of bilateral linear hearing aids have been fit thus far. Current efforts are devoted to defining the signal-processing algorithms for laboratory testing.

1.7 Binaural Hearing

Sponsor

National Institutes of Health
Grant R01 DC00100²⁹

Project Staff

Nathaniel I. Durlach, Dr. Kourosh Saberi, Dr. Patrick M. Zurek

1.7.1 Overview

The long-term goal of this project is: (1) to develop an integrated, quantitative theory of binaural interaction that is consistent with psychophysical and physiological data on normal and impaired auditory systems, and (2) to apply our results to the diagnosis and treatment of hearing impairments. Current work is in two areas: (1) a study of the lateralization of transient noise bursts with time-varying interaural delays, and (2) analysis of the acoustic cues that underlie the localization of "phantom" sound images in intensity-difference stereophony.

1.8 Virtual Environment Technology for Training: Enabling Research on the Human Operator

Sponsor

U.S. Navy - Office of Naval Research/
Naval Air Warfare Center
Contract N61339-95-K-0014
Contract N61339-96-K-0003

Project Staff

Nathaniel I. Durlach, Dr. Mandayam A. Srinivasan, Dr. Thomas E.v. Wiegand, Dr. G. Lee Beauregard, Dr. Cagatay Basdogan, David W. Schloerb, Walter A. Aviles, Adrienne H. Slaughter, Kari Anne H. Kjolaas, Lajos Molnar, Evan F. Wies, Chih-Hao Ho, I-Chun A. Hou, David E. DiFranco

1.8.1 Program Summary

This work is being conducted within Virtual Environment Technology for Training (VETT), a large interdisciplinary, inter-institutional program which is studying the use of virtual environment (VE) technology to improve Navy training. The Enabling Research on the Human Operator (ERHO) component of the VETT program is concerned with: (1) how human perception and performance in virtual environments (VEs) depend upon the physical characteristics of the VE system, (2) the task being performed, and (3) the user's experience with the system and the task. To the extent that the ERHO research is successful, the results will not only provide important information for the design and

²⁹ Subcontract from Boston University. Professor H. Steven Colburn is the principal investigator.

evaluation of VE training systems, but also for VE systems in general.

The work performed in this program involved continuation of experimental research in the following five areas. In section 1.8.2, "Human Responses to Alterations in Sensorimotor Loops," studies were conducted to determine and model the human user's ability to adapt to distorted relationships between seen (visual) and felt (proprioceptive/ kinesthetic) hand position. In section 1.8.3, "Haptic Psychophysics," research continued on the human user's ability to discriminate and identify the mechanical object properties of compliance, viscosity, and mass. In "Depth Perception in VE Systems," section 1.8.4, we are exploring limitations both on the use of stereopsis and the use of motion parallax to provide accurate depth information. In section 1.8.5, "Multimodal Interactions with Force Feedback," we are determining interactive effects among the visual, auditory, and haptic modalities. In section 1.8.6, "Sensorimotor Involvement for Training Cognitive Skills," we are evaluating the benefits of such involvement, particularly haptic actions, in simple memory tasks.

1.8.2 Human Responses to Alterations in Sensorimotor Loops

VE systems generally cause alterations in sensorimotor loops because they do not replicate reality perfectly. These deviations from reality may result from unavoidable equipment limitations, trade-offs that have been made to reduce cost, or distortions that are introduced intentionally to enhance performance. Independent of the cause of the alteration, to improve our ability to design cost-effective VE systems, it is necessary to predict human responses to such alterations as a function of the alteration considered and of the user's experience with this alteration.

During the past year, we continued a series of experiments concerned with altered relations between seen and felt hand position. The experimental task used in these studies has been described in previous RLE Progress Reports. Briefly, the task used, the radial Fitts task, consists of repeatedly moving a cursor (controlled by hand via a manipulandum) from a central "home" location to a random target location. Targets appear at locations described by a presentation radius and a presentation angle, in a coordinate system centered around the home point. During home-to-target motion, the cursor is not visible (open loop). When the manipulandum stops moving, the trial ends and the cursor reappears, allowing the subject to visu-

ally compare the final cursor location with the target's position (visual feedback).

Figure 1 shows a flow chart of the aspects of the sensorimotor loop that we wish to model. The "virtual task" is the algorithm that positions the random targets according to the scripted experimental procedure. The alteration function, f , modifies the signal from the manipulandum, which positions the cursor on the display screen. The compensation function, g , describes the mapping from the position of the target seen by eye to the hand motion used to guide the cursor to the target position. The compensation function adapts to the alteration function based on the error in cursor position which is viewed at the end of each trial. The details of how the series of end-of-trial error signals cause g to change from the null function ($g(x)=x$) to a function that behaves more like the inverse of the alteration function ($g(x)=f^{-1}(x)$) are not covered in the current model. Instead, we explore a number of likely compensation functions that appear to describe the empirical data collected under the effects of a small number of alteration functions. These particular alteration functions have been chosen according to tentative theories of the space of possible compensation functions and, more generally, the developmental nature of the eye-brain-hand wetware that forms the important half of the loop.

Two of the experimental alteration conditions that we have used include independent manipulandum gain on x and y axes, i.e.:

$$(x_{\text{cursor}}, y_{\text{cursor}}) = (x_{\text{gain}} \cdot x_{\text{manip}}, y_{\text{gain}} \cdot y_{\text{manip}})$$

and a more general interaction condition specifically designed to test the emerging model, having the functional form:

$$(x_{\text{cursor}}, y_{\text{cursor}}) = (f_x(x_{\text{manip}}, y_{\text{manip}}), f_y(x_{\text{manip}}, y_{\text{manip}})).$$

The current model is based on an aggregate of two classes of adaptation behavior observed in the data. We call these classes of adaptation the x,y model and the r,θ model as summarized below.

Adaptation Model: Perfectly Compensated Radius and Angle

Perfectly compensated radius and angle model adaptation is prescient selection of the inverse of the alteration function for use as the compensation function (in rectangular coordinates), the x,y model. In the PCRA model, the compensation function g is simply the inverse of f , multiplied by an adaptation factor $0 \leq k \leq 1$ (k will usually be an exponential

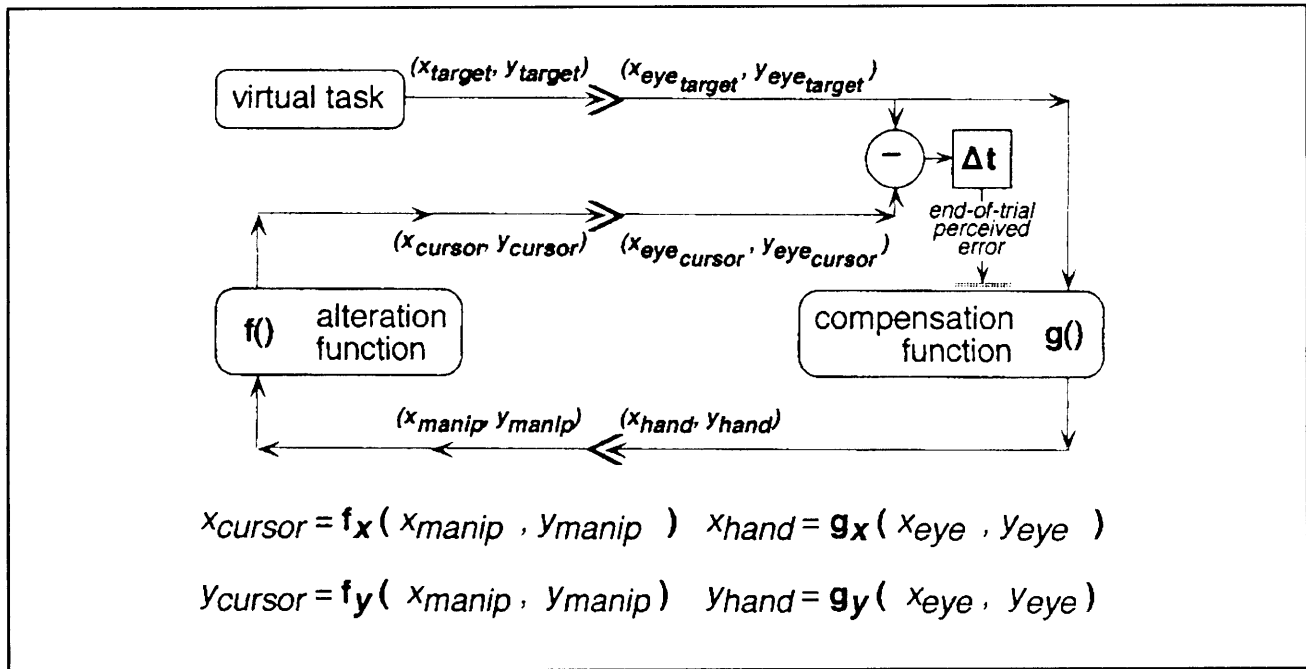


Figure 1.

function of the duration of experience, e.g., $k = 1 - e^{-\text{trials}/\text{rate}}$. The equation describing the error at the end of each trial is:

$$(x_{\text{error}}, y_{\text{error}}) = (r_{\text{target}} \cdot \cos(\theta_{\text{target}}) \cdot (1 - k), r_{\text{target}} \cdot \sin(\theta_{\text{target}}) \cdot (1 - k))$$

(as derived in our paper *in vivo*). The most salient feature of this model is that as k increases asymptotically to 1, the error decreases to zero, thus predicting perfect learning. Although perfect learning is not observed in the data, the model does appear to match the data during the earlier part of the adaptation process, most notably in the pattern of decrease of angular error.

Adaptation Model: Average Radius, Same Angle

The adaptation model minimizes average radial error, without adapting to angular error, (the "r, θ " model). In the ARSA model, the compensation function g is derived from a proposed simple heuristic that is used by the subject to minimize radial error without compensating for angular error. The equation describing the error at the end of each trial in this case is:

$$(x_{\text{error}}) = x_{\text{gain}}(k - 1) + (r_{\text{target}} - k \bar{r} \cdot x_{\text{gain}}) \cdot \cos(\theta_{\text{target}})$$

$$(y_{\text{error}}) = y_{\text{gain}}(k - 1) + (r_{\text{target}} - k \bar{r} \cdot y_{\text{gain}}) \cdot \sin(\theta_{\text{target}})$$

(derived in our paper *in vivo*). This model, predicts the correct pattern of radial error with increased adaptation; however it does not predict any improvement in angular error.

Examination of a model in which a two-stage process, incorporating a fast process (short-time constant) based on the PCRA model, as well as a secondary slow process based on the ARSA model, appears to account for the major features of the data we have collected. To further challenge this combined model, we have run a set of conditions in which there is no straightforward functional description in either a cartesian or a polar coordinate system. These conditions instead follow a rule similar to the process proposed in the ARSA model. In this model, the radial position of the displayed cursor is always the same as that of the manipulandum, but the displayed angle is scaled. We are still analyzing the data from this set of experiments and will include these results in an upcoming paper.

In addition to the work described above on adaptation to altered relations between seen and felt hand position, we have completed a series of three articles for publication on the work previously done in the area of altered auditory localization cues. See section 1.12, "Research on Supraauditory Localization for Improved Human-Machine Interfaces."

1.8.3 Haptic Psychophysics

Concurrent with the technology development that enables one to realize a wider variety of haptic interfaces, it is necessary to characterize, understand, and model the basic psychophysical behavior of the human haptic system. Without appropriate knowledge in this area, it is impossible to determine specifications for the design of cost-effective haptic interfaces. Human abilities and limitations play an important role in determining the design specifications for the hardware and software that enable haptic interactions in VE. With this viewpoint, psychophysical experiments have been carried out with a haptic interface to measure human haptic resolution in discriminating fundamental physical properties of objects through active touch. The linear grasper, a computer controlled electromechanical apparatus, was developed and used in these experiments. The subjects utilized their thumb and index fingers to grasp and squeeze two plates of the linear grasper, which was programmed to simulate various values of the stiffness, viscosity, or mass of virtual objects. During the experiments, haptic motor performance data in terms of applied forces, velocities, and accelerations were simultaneously recorded.

The just noticeable difference (JND), a commonly accepted measure of human sensory resolution, was found to be about 7 percent for stiffness, 12 percent for viscosity, and 20 percent for mass. The motor data indicated that subjects used the same motor strategy when discriminating any of these material properties. Further analysis of the results has led to the postulation of a single sensorimotor strategy capable of explaining both the sensory resolution results and motor performance data obtained in the experiments. The temporal force control-spatial force discrimination (TFC-SFD) hypothesis states that subjects apply the same temporal profile of forces to all stimuli and discriminate physical object properties on the basis of differences in the resulting spatial profiles of these forces. Implications of these results in the design of virtual environments include specifications on how accurately the dynamics of virtual objects need

to be simulated and what parameter values will ensure objects that can be discriminated.

Research in this VETT supported project has resulted in a doctoral dissertation and several papers.³⁰ Two more journal papers are being prepared.³¹

1.8.4 Depth Perception in Virtual Environment Systems

One of the supposed (and highly acclaimed) advantages of VE systems is the three-dimensional nature of the visual images (specifically, the perception of depth in these images). Limitations on depth perception in VE systems are likely to arise, however, because of limited screen resolution (non-infinitesimal pixels) and limited temporal sampling (finite frame rate). Whereas the limited spatial resolution is likely to degrade depth perception via stereoscopy, the limited temporal sampling is likely to degrade depth perception via motion parallax. Obviously, knowledge about these degradations and how they can best be overcome is important for effective VE design.

Depth Perception via Stereoscopy

We have completed an extensive series of experiments on stereoscopic depth perception in VEs that provides data both on the degradations associated with finite screen resolution (pixelization) and on the enhancements that can be achieved by simulating increased interocular distance. The results obtained in this research reinforce the notion that (1) stereoscopic depth perception is significantly influenced by pixelization, and (2) one can readily adapt to an increase in interocular distance by at least a factor of two. In general, it appears that the primary limitations on the benefits in depth perception that can be achieved by expanding the interocular distance relate less to constraints on the ability to adapt to such an expansion than to the physical limitations of the display (pixelization) and to the psychophysical limitations inherent in absolute identification of any stimulus parameter

³⁰ H.Z. Tan, N.I. Durlach, G.L. Beauregard, and M.A. Srinivasan, "Manual Discrimination of Compliance Using Active Pinch Grasp: The Role of Force and Work Cues," *Percept. Psychophys.* 54: 4, 495-510 (1995); G.L. Beauregard, M.A. Srinivasan, and N.I. Durlach, "Manual Resolution of Viscosity and Mass," *Proceedings of the ASME Dynamic Systems and Control Division*, DSC-Vol. 57: 2, 657-662 (ASME, 1995); G.L. Beauregard, "Sensorimotor Interactions in the Haptic Perception of Virtual Objects," Ph.D. diss., Dept. of Biomed. Eng., Boston University, 1996; G.L. Beauregard and M.A. Srinivasan, "Sensorimotor Interactions in the Haptic Perception of Virtual Objects," *The Engineering Foundation Conference on Biomechanics and Neural Control of Movement*, Mt. Sterling, Ohio, June 1996.

³¹ G.L. Beauregard, M.A. Srinivasan, and N.I. Durlach, "Manual Discrimination of Viscosity and Mass through Active Pinch Grasp," in preparation; G.L. Beauregard and M.A. Srinivasan, "Sensorimotor Interactions in Haptic Perception: The Role of Force Variations in Time and Space," in preparation.

(including depth) when the range of variation of the parameter to be identified is large relative to the just-noticeable difference in that parameter. David Schloerb's Ph.D. dissertation describes in detail the facilities used for this research, the experimental paradigms explored, the formal data collected, the analysis techniques exploited, and the manner in which the results of these experiments relate to various theoretical models that we have developed in the areas of absolute identification and sensorimotor adaptation.³²

Depth Perception via Motion Parallax

Motion parallax is compelling due to depth in situations in which the viewer is able to freely move, such as occurs with HMD-based VEs. Thus the manner in which motion-parallax mediated depth perception is affected by imperfect temporal and spatial characteristics of the VE interface is an important topic for study.

To support the motion parallax experiments, we have constructed a device to passively constrain an observer's head motion according to selectable velocity profiles. This system is constructed around a standard "bite bar" which can slide freely along a 20 cm track. In the experiments, the observer's self-generated lateral motion can be limited through the frictional drag imparted by a servo-controlled magnetic particle brake. We are thus able to safely achieve changes in the velocity profile of the observer's viewpoint, while completely avoiding components (i.e., motors) that are capable of generating motion on their own and possibly injuring the subject.

In the current experiments, subjects view a split-field random dot display, and we measure their ability to identify the nearer field. The data obtained thus far suggest that degradation of depth perception is relatively small for delays of less than 170 msec; however, a precipitous increase in discrimination threshold occurs when delay increases beyond 200 msec.

Unfortunately, unexpected problems with the reliability of the software (i.e., frequent system crashes) and the loss of the primary researcher on the project have delayed our progress. During the months since the departure of the primary researcher, we have solved the software problems, and are now beginning to resume the running of experiments.

1.8.5 Multimodal Interactions with Force Feedback

Because multimodal sensorimotor involvement constitutes a key feature of VE systems, it is obviously important to understand multimodal interactions. Furthermore, because the availability of force feedback in multimodal VE interfaces is relatively new, knowledge about interactions involving force feedback is relatively limited. In general, research in this area not only provides important background for VE design, but the availability in VE systems of multimodal interfaces with force feedback provides a unique opportunity to study multimodal sensorimotor interactions.

To explore the possibility that multisensory information may be useful in expanding the range and quality of haptic experience in virtual environments, experiments have been conducted to assess the influence of visual and auditory information on the perception of object stiffness through a haptic interface. We have shown that visual sensing of object deformation dominates kinesthetic sense of hand position and results in a dramatic misperception of object stiffness when the visual display is intentionally skewed.³³ However, the influence of contact sounds on the perception of object stiffness is not as dramatic when tapping virtual objects through a haptic interface.³⁴ In the following paragraphs, we first summarize the visual-haptic interaction experiments, then the auditory-haptic interaction experiments.

³² D. Schloerb, *Perception of Depth via Stereoscopy in Virtual Environments: Effects of Increasing Interocular Distance*. Ph.D. diss., Dept. of Mech. Eng., MIT, 1997.

³³ M.A. Srinivasan, G.L. Beauregard, and D.L. Brock, "The Impact of Visual Information on Haptic Perception of Stiffness in Virtual Environments," *Proceedings of the ASME Dynamic Systems and Control Division*, DSC-Vol. 58: 555-559, ASME (1996).

³⁴ D.E. DiFranco, G.L. Beauregard, and M.A. Srinivasan, "The Effect of Auditory Cues on the Haptic Perception of Stiffness in Virtual Environments," submitted to the Winter Annual Meeting of the ASME.

Visual-Haptic Interactions Involving Force Feedback

The impact of visually presented spatial cues on the human perception of mechanical stiffness of virtual objects was investigated using the planar grasper, a three-degree of freedom, force-reflecting haptic interface developed at MIT. In each trial, the subjects manually pressed and felt the stiffness of two springs programmed to be side-by-side within the workspace of the planar grasper. At the same time, visual images of the springs were displayed on a computer monitor so that the amount of deformation of the spring that was being pressed was displayed in real-time. The subjects were asked to judge which of the two springs was stiffer. Unknown to the subjects, the relationship between the visually presented deformation of each spring and the actual haptic deformation was systematically varied between experimental trials. This relationship ranged from fully registered (visual deformation equal to the actual deformation of each spring) to completely interchanged (visual deformation of the softer spring equal to the haptic deformation of the harder spring and vice versa).

The results demonstrated a clear dominance of the visual sense over the kinesthetic sense for hand position. The subjects essentially ignored all kinesthetic hand position information regarding spring deformation and based their judgment of stiffness on the relationship between the visual position information and the indentation force sensed tactually. Both in the absence of visual information (when the monitor was turned off), as well as when the visual displacement was exactly equal to the haptic displacement, the subjects were almost 100 percent correct in their judgments. However, with increasing mismatch between the visual and haptic position information, an increasing misperception of stiffness occurred, culminating in totally erroneous judgments when the two were interchanged. These and other experiments involving different scaling values between the visual and haptic displacements showed that in computing the stiffness of objects, when subjects have a choice of associating visual or haptic displacements with their haptic perception of force, they consistently choose the visual displacements, i.e., the visual information dominates the kinesthetic information in this context.

An important implication of these results for multimodal VEs is that by skewing the relationship between the haptic and visual displays, the range of object properties that can be effectively conveyed to the user can be significantly enhanced. For example, although the range of object stiffness that can be displayed by a haptic interface is limited by the force-bandwidth of the interface, the range per-

ceived by the subject can be effectively increased by reducing the visual deformation of the object.

Auditory-Haptic Interactions Involving Force Feedback

Psychophysical experiments were conducted to assess the influence of contact sounds on the perception of object stiffness during tapping virtual objects. The PHANToM, a haptic interface recently developed at MIT with partial support from the VETT project, was used to display the virtual objects to the subjects. The PHANToM has six degrees of motion freedom together with force-reflection along three axes. As subjects tapped on the different virtual surfaces, they were simultaneously presented with previously recorded impact sounds such as the sound of a stylus striking cloth, wood, or metal. The subjects were asked to rank the surfaces based on their perceived stiffness.

The results showed that when the physical stiffness of the surfaces was the same, subjects consistently ranked the surfaces according to sound, i.e., surfaces paired with sound cues that are typically associated with tapping harder surfaces were generally perceived to be stiffer. However, when sound cues were randomly paired with surfaces of different mechanical stiffnesses, the results were more equivocal: naive subjects who had not used the PHANToM previously tended to be more affected by sound cues than another group of subjects who had previously completed a set of stiffness ranking experiments without sound cues. It appears that the effect of sound cues, although significant, is not as dominant as that of visual cues (described above) in influencing the haptic experience of object stiffness.

1.8.6 Sensorimotor Involvement for the Training of Cognitive Skills

As indicated previously, an important component of the justification for the added expense associated with immersive, multimodal, interactive VE systems is the idea that sensorimotor involvement facilitates the training of cognitive skills. Accordingly, it is necessary to determine the extent to which, and the conditions under which, this idea is correct, and to develop a model that is consistent with these findings and can be used to make reliable predictions.

Work to date on this project has focused on the effects of such involvement (particularly haptic) in a memory task in which the subject is required to memorize short sequences of letter-number pairs. This particular memorization task (i.e., memoriza-

tion of short sequences) was chosen because it is simple enough to be experimentally and theoretically tractable and yet, at the same time, representative of real-world control panel tasks in which the user is required to identify the state of the system to be acted upon, and then to recall the series of actions that should be taken when the system is in that state.

In these experiments, the stimulus set consists of eight sequences of length five each. The first element (SD, SX, DB, or SF) designates the state of the system, and the next four elements (drawn from the set A1, A2, A3, B1, B2, B3, C1, C2, C3), the series of actions to be taken for that state. In each trial for each experiment, the subject was presented with a state element and then required to specify the corresponding sequence of four action elements. In general, an experiment consisted of a practice session in which the subject attempted to learn the sequences and a testing session in which the ability to recall the sequences was estimated. Testing sessions were performed five minutes, 30 minutes, and 24 hours after the practice session in order to explore the time dependence of recall performance.

Two sets of sequences were presented to the subjects in the practice session. Subjects in the control group learned both set 1 and set 2 under the control condition. Subjects in the experimental group learned set 1 under the control condition and set 2 under the experimental condition. This design enabled us to look at results both within an individual and across populations of individuals.

In experiment 1 (in which 20 subjects took part), the control condition involved memorizing the sequences using only a visual string presentation (e.g., SZ, A1, C2, B2, C3). In contrast, in the experimental condition, the subject was also provided with a visual graphic of the 3 X 3 matrix displaying all the action elements (rows A, B, C; columns 1, 2, 3) and was required to trace the sequence of actions to be memorized by moving a finger around the matrix in the appropriate manner (e.g., for the above sequence, from A1 to C2 to B2 to C3). Thus, for the experimental group in the experimental condition (set 2), the sequence was associated with a visual and haptic image of the geometry of the sequence (i.e., the locations of the elements and the lines joining successive elements in the sequence). Although there was substantial intersubject variability, performance was clearly better in the experimental condition for all testing delay times: in all cases, the average test score for the experimental group in the experimental condition (set 2) was roughly twice that of the score for the experimental group in the control condition (set 1) or the score for the control group (in either set 1 or set 2).

In experiment 2 (which again made use of 20 subjects), the same control and experimental conditions were used except that the control condition included the visual (not the haptic) representation of the matrix used in the experimental condition (Set 2 of the experimental Group) of experiment 1.

The results of experiment 2, unlike those of experiment 1, were puzzling. On one hand, the results for the experimental group showed a decided advantage for the experimental condition (visual plus haptic matrix) over the control condition (visual matrix only), thus suggesting that haptic tracing of the matrix improved performance over merely visual sensing of the matrix. On the other hand, many questions are raised by the following results of experiment 2: the scores of the control group in both sets 1 and 2 are much greater than the scores of the experimental group in the control condition (set 1) and are roughly comparable to the scores of the experimental group in the experimental condition (set 2).

A detailed description of most of this work to date on this project is available in the Masters thesis of Walter A. Aviles, (see below).

1.8.7 Publications

Aviles, W.A. *The Role of Sensorimotor Involvement in the Learning of Cognitive Skills*. S.M. thesis, Dept. of Electr. Eng. and Comput. Sci., MIT, 1996.

Beauregard, G.L. *Sensorimotor Interactions in the Haptic Perception of Virtual Objects*. Ph.D. diss., Dept. of Biomed. Eng., Boston University, 1996.

Beauregard, G.L., and M.A. Srinivasan. "Sensorimotor Interactions in the Haptic Perception of Virtual Objects." *The Engineering Foundation Conference on Biomechanics and Neural Control of Movement*, Mt. Sterling, Ohio, June 1996.

Beauregard, G.L., M.A. Srinivasan, and N.I. Durlach. "Manual Discrimination of Viscosity and Mass through Active Pinch Grasp." In preparation.

Beauregard, G.L., and M.A. Srinivasan. "Sensorimotor Interactions in Haptic Perception: The Role of Force Variations in Time and Space." In preparation.

DiFranco, D.E., G.L. Beauregard, and M.A. Srinivasan. "The Effect of Auditory Cues on the

Haptic Perception of Stiffness in Virtual Environments." Submitted to the Annual Winter Meeting of the ASME.

Schloerb, D. *Perception of Depth via Stereoscopia in Virtual Environments: Effects of Increasing Interocular Distance*. Ph.D. diss., Dept. of Mech. Eng., MIT, 1997.

Srinivasan, M.A., G.L. Beauregard, and D.L. Brock. "The Impact of Visual Information on Haptic Perception of Stiffness in Virtual Environments." *Proceedings of the ASME Dynamic Systems and Control Division*. DSC-Vol. 58: 555-559, ASME (1996).

1.9 Spatial Knowledge Acquisition and Training Using Virtual Environments

Sponsor

U.S. Navy - Office of Naval Research
Grant N00014-96-1-0379

Project Staff

Nathaniel I. Durlach, Dr. Thomas E.v. Wiegand,
Glenn Koh, Maciej Stachowiak, Bryan D. Kincy

1.9.1 Introduction

The overall objective of this work, which is being conducted in collaboration with Rudy Darken at the Naval Postgraduate School, is to apply state-of-the-art virtual reality simulation technologies to the investigation of issues in spatial knowledge acquisition and training. Virtual environment training provides a number of advantages over traditional training methods and it is expected that such training will translate into measurable real-world gains in performance. The first set of studies are intended to explore the extent to which virtual environments are an effective means of conveying spatial knowledge of a specific real space. The second set, which will not be initiated until later on in the program, will be concerned with the training of spatial knowledge and navigational skills in

general and with the ability to transform spatial information presented in one form to information presented in other forms. Throughout all of our work, we will explore the role of various environmental features and stimulus characteristics in navigation and wayfinding so that we determine which components of virtual-environment fidelity can be degraded without causing substantial reductions in training transfer. In this manner, we hope to develop systems that are optimally cost effective. To the extent possible, we will also attempt to shape our research to the needs of special communities (such as special operations groups) for whom detailed accurate spatial knowledge of spaces that are not accessible for training is of great practical importance. Because the work involves both basic and applied research, and is also highly interdisciplinary (it involves cognitive science, perceptual psychology, virtual-environment design, real-world training methods, etc.), it requires close collaboration among a wide variety of individuals. General background on the use of virtual environments for spatial training can be found in Darken, 1996, and in the references cited there.³⁵

The first year of work on this project has been characterized by concern with the acquisition of spatial knowledge for specific spaces and with the development of facilities and procedures for conducting experimental work. Furthermore, attention has been focused on the acquisition of configurational knowledge as opposed to route knowledge. Whereas route knowledge is characterized by a navigator's ability to move from one position to another along a specific route, configurational knowledge is characterized by information on the relationships between different locations and the overall structure of the space, as well as by the ability to assume exocentric viewpoints. A navigator with strong configurational knowledge can find optimum routes between two points in the environment, draw maps of the environment, and estimate the direction and distance to a particular landmark from an arbitrary position in the environment. We have focused on configurational knowledge rather than route knowledge because it is fundamentally more important and because the use of virtual environments for providing route knowledge has been more extensively studied in the past.³⁶

³⁵ R.P. Darken, "Wayfinding in Large-Scale Virtual Worlds," Ph.D. diss., School of Engineering and Applied Science, George Washington University, Washington, D.C., 1996.

³⁶ B.G. Witmer, J.H. Bailey, and B.W. Knerr, "Training Dismounted Soldiers in Virtual Environments: Route Learning and Transfer," U.S. Army Research Institute for the Behavioral and Social Sciences, 1995.

1.9.2 Progress Report

Our work during the past year can be summarized under the following headings: (1) experimental design, (2) VE implementation, and (3) support efforts.

Experimental Design

Task Selection

Configurational knowledge will be tested using two methods, each of which controls for external factors such as verbal descriptive abilities: (1) a *pointing* task in which the subject is brought to a position A and asked to estimate direction and distance to a number of other positions B in the venue, and (2) a *wayfaring* task in which the subject is asked to proceed from location A to location B via some route.

The *pointing* task measures the subject's ability to reproduce a 3D replica of the space by having him point and estimate the distance to landmark B from position A (which may not be the subject's current position). Theoretically, a subject who is able to identify the locations of all points relative to his own or relative to each other has a perfect spatial knowledge of the venue. The accuracy (in azimuth and distance) of the subject's responses can be used to evaluate the degree of knowledge of the space.

The *wayfaring* task involves having the subject move between two points in the venue via a route with characteristics described by the interviewer at the time of testing. The subject may be asked to find the shortest path between A and B or a path via a number of waypoints. The goal of this test is to assess the subject's spatial and navigational knowledge from an egocentric perspective. Because the test paths are not known to the subject during the practice phase of the experiment, the subject is forced to piece together observations of the venue derived from information obtained during the practice phase—it is not possible for the subject to memorize the specific routes to be tested. The subject's performance in the wayfaring task will be assessed using a number of variables, including: (1) time taken to complete the task, (2) number of wrong turns, (3) distance travelled by the subject, and (4) ability to complete the task.

Venue Selection

The selection of an experimental site is clearly of great importance to continued success of the

project. The site must be of sufficient complexity to reveal differences in the experimental conditions in current and subsequent studies. Because creating a virtual environment is time-consuming, it is important to choose a site which will be available and accessible for a period of time and also to create a model with sufficient detail to be useful for further experiments.

The benchmark for the virtual environment model is to produce the most accurate model and to create the deepest sense of immersion possible within the limits of VE modeling. Starting with a high-resolution model, it is possible to degrade the model in different respects as experimental conditions. Starting with a specific, less detailed, model would require the creation of a new model for each new VE feature introduced.

The goal, then, has been to select a venue with substantial complexity and to create a virtual model with sufficient detail that the VE model and real-world venue pair can be used for the next several years as an experimental system without complex modifications of the system (thereby amortizing the initial high modeling time-investment over a span of many experiments).

The venue chosen for modeling was MIT Building NW30. This building is a "dead storage" warehouse and was chosen because of its relative complexity, availability, and access for experimentation with large numbers of subjects. NW30 is on the MIT campus and approximately a 15-minute walk from our main laboratory at RLE (where the VE equipment is located). An additional appealing factor supporting the use of NW30 is that it is generally not visited by students, and hence few of our potential subjects will be disqualified from the study due to previous experience in the test venue.

The model comprises four of the five levels of NW30. The basement of NW30 was not included in the model because the four above-ground levels provide sufficient complexity for the study. The NW30 building spans approximately 89,695 square feet. As indicated in figure 2, a large number of loops and decision points are available. Loops are conduits which connect to themselves—if a subject follows a loop, the subject will eventually return to the starting point of the loop. Decision points are defined as branches in a path (or conduit) with two or more possible choices of direction. The complexity of a given venue is indicated by the number of decision points and loops in that venue. In table 1, we summarize the single-floor-only loops and decision points that are easily usable in the chosen venue.

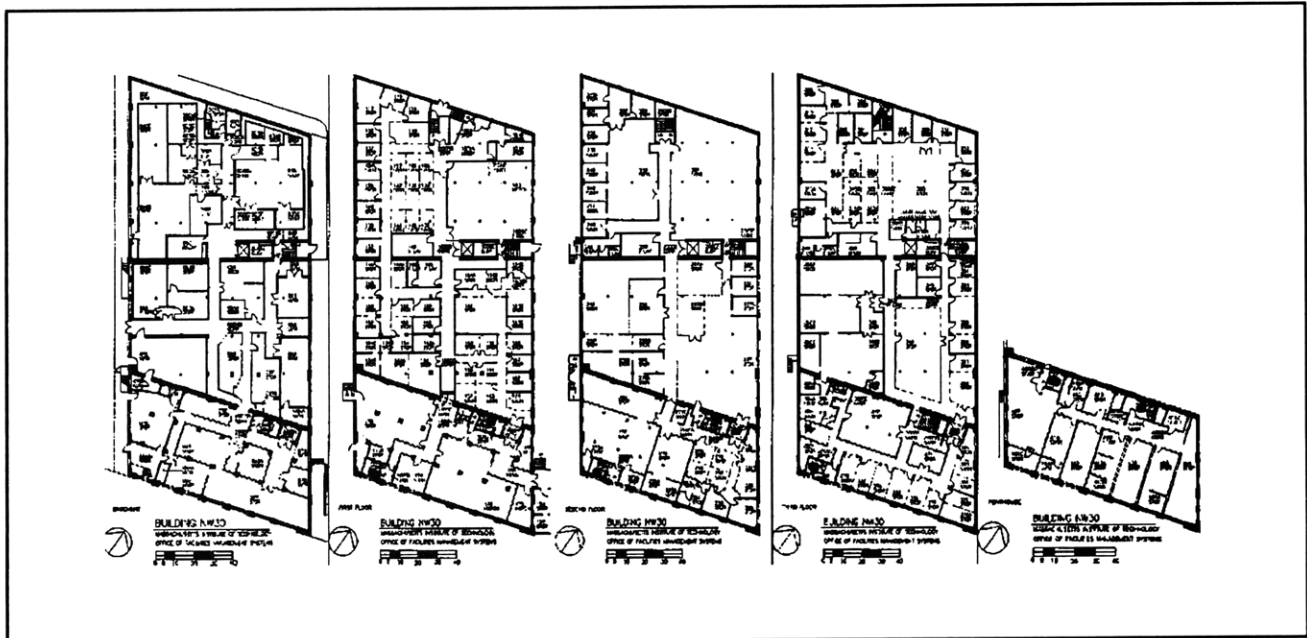


Figure 2. Floor Plan of MIT Building NW30.

Table 1: Single-floor decision points and loops in NW30.

Building	Decision Points	Loops
Level	(not including stairwell)	(2D only)
Floor 1	12	3
Floor 2	4	0
Floor 3	6	1
Floor 4	1	0

Note that the selected venue contains a number of stairwells, and that each stairwell is also a decision point. The effective complexity of the building is thus multiplied because of the many loops which exist once the stairwells and the various 3D paths are taken into consideration. Additionally, the building contains an elevator and a mezzanine level which is generally not accessible; these features can be pressed into service if future experiments require the added complexity.

The selected building and its associated virtual model will allow for many subsequent experiments because of its complexity and multilevel nature. Complexity can be arbitrarily increased by selecting paths with multistory waypoints. Conversely, it is possible to use the virtual model for a less complex navigational study simply by limiting the subject to one floor.

Training Phase

There are four training conditions to be used in connection with the above described venue:

1. Traditional training using maps only.
2. Exploration of venue with maps.
3. Exploration of virtual reality representation of venue with maps.
4. Exploration of virtual reality representation of venue with maps and virtual reality training aids.

Subjects are allocated evenly to these four groups. Subjects have no prior knowledge or experience with the experimental venue. The experiment proceeds in two phases: the initial training phase utilizing one of the above four treatments, and a testing phase, elaborated further below.

The training phase allows subjects to acquire some form of knowledge about the venue. For each of the training conditions, the subjects are given a fixed amount of time during the training to digest and acquire some knowledge of the space. All subjects are given 20 minutes of training and are encouraged to spend the entire allotted time learning the environment.

Training involves giving the subject a map of the venue with certain landmarks and locations highlighted. All four groups of subjects receive identical maps as a learning tool. The subjects are notified that the highlighted points on the map are of partic-

ular importance in the task. Subjects in training conditions 2, 3, and 4 are allowed to explore some representation of the space charted by the map (the actual building for condition 2, and the virtual model for conditions 3 and 4). Subjects are given free exploration of the venue in both the virtual environment and the real world conditions with no input from the experimenter, aside from the suggestion that the subject visit all the highlighted locations. The only constraint on exploration is the venue itself (available conduits in both the real and virtual environments and the 20 minute time limit).

Subjects in conditions 3 and 4 are exposed to the virtual environment training system. They are given as much time as is necessary to familiarize themselves with the VE and the interface equipment. The subjects in these two groups are allowed to gain knowledge of the system and its navigational tools by exploring an unrelated "sample" space in the VE (consisting of a model of a part of our main laboratory). In this way, the subject will be able to learn to be comfortable operating in the VE, and will gain familiarity with the system without gaining extra time to learn about the experimental venue.

Two forms of the VE are available, assigned to the two VE training groups:

1. The first VE is the standard high-resolution model.
2. The second VE is enhanced to allow the subject to select "translucent walls on demand" in effect giving the subject a sort of "x-ray" vision.

Subjects in all four groups are given 20 minutes to explore the environment. At the end of the 20 minutes, subjects proceed to the actual venue for the testing phase.

Testing Phase

Subjects from the four treatment groups are tested in the actual venue. All subjects are presented with the same set of tasks; in both the pointing and the navigation tasks, the subjects are asked to identify the same locations and navigate the same routes.

The testing phase consists of six tasks: three pointing tasks and three navigational tasks. The tasks are interleaved, with subjects alternating

between the navigational and the pointing tasks. Upon entering the experiment site, subjects are asked to walk to a position A from the entrance of the warehouse using the shortest possible route. At this target point, the subject is asked to point and estimate the distance to a location B. This process is repeated twice more, with the subject moving from the current position to a new position through a route described by the experimenter. After this cycle of three sets of two tests, the testing phase for that particular subject is complete.

The three pointing tests require the subject to:

1. Point at locations on the same level.
2. Point to locations on a different level.
3. Point at a location at an initial point displaced from his own. The subject will be asked to place himself in the frame of reference of a remote location and point toward a different remote location.

The wayfinding tasks consist of:

1. Finding the shortest path to a given location.
2. Finding a path via a number of waypoints.
3. Finding a path to a location when an unexpected barrier is introduced.

Because the target points are identified in the real world venue, the subject should always be able to eventually reach the next location as identified by the interviewer. If the subject can not do so within a reasonable amount of time or seems lost (as determined by the experimenter), the subject is brought to the next location and asked to complete the pointing task at that location. Mistakes made by the subject are not identified by the interviewer to the subject. In tasks in which the subjects are expected to traverse a route via some set of waypoints, missing a waypoint should be obvious to the subject as the waypoints are easily identifiable locations or objects (water-fountains, fire extinguishers, etc.).

All subjects are evaluated with the same criteria independent of the training method. The tabulated data of table 2 are then analyzed to determine the relationships between task performance and training method.

Table 2: Test-phase measurements for each subject.

Navigational Task	Time Taken	Distance Traveled	Number of Wrong Turns	Number of Missed Waypoints
Wayfinding 1				
Wayfinding 2				
Wayfinding 3				
Pointing Task	Azimuth	Distance		
Pointing 1				
Pointing 2				
Pointing 3				

VE Implementation

Overview

Providing a head tracking viewpoint in a virtual environment is a complex problem. While equipment exists to track position and orientation, and to display stereographic images on a head-mounted display, integrating the separate components with existing simulation and rendering software is still awkward. We have integrated available components to provide a "walk-through" simulation that allows a subject to look around and have his point of view in the simulation follow the angle of his head. Furthermore a joystick allows two different modes of control for walking around, and collision detection prevents the subject from walking through walls.

Both the input/output devices and a simulation software package to render the building model were already available. The simulated environment itself was developed using an existing software package on a Silicon Graphics workstation. The available input-output device is a head-mounted stereoscopic display with an attached 6D position/rotation sensor. The main focus of the project was integration of available tools, in particular to receive 6 DOF data from the tracker and use it to control the simulation engine and render to the head-mounted display.

The building model was developed with Designers Workbench (DWB) from Coryphaeus Software, a modeling tool which allows for the creation of 3D models. DWB facilitated the building of the model with its built in texture mapping capabilities. EasyScene, a software package also from Coryphaeus Software, generates the virtual world and allows for the recording of the subject's performance in the virtual world. EasyScene allows for full tracking of the subjects' actions, including the

distance travelled, time spent moving, direction the subjects look at, etc.

This study uses a 2 processor Onyx with the Reality Engine II graphics processor. A stereoscopic head-mounted display (HMD) is used as the subjects' visual interface in the virtual environment training condition. This HMD is keyed to the directionality of the subjects head, tracked by a Polhemus system. Control devices for the subject include (1) Mouse and Keyboard, (2) Slippery Floor, Slippery Fingers, and (3) Magic Wand. The experimenter's workstation is equipped with a computer monitor which provides a real-time 2D image of what the subject sees.

Texture maps are added to the DWB model in order to create a photorealistic virtual representation of the real-world venue. Collecting texture maps on a large scale is non-trivial, and traditional methods of digitizing photographic images become extremely time-consuming in such a large venue. An automatic room scanner will be used to facilitate the gathering of texture data.

Implementation of Serial Input Devices

Implementing the head tracking control system required careful attention to several details. Interfaces to the hardware devices had to be implemented. This required handling of the RS-232C protocol over a serial line, as well as handling of the device-specific protocols for the joystick and the tracker. One module implemented the RS-232C serial interface. Under the workstation's operating system, serial channels are typically expected to be interactive terminals, so much preprocessing of the input stream is done by the operating system itself by default. However, the devices used are not much like a terminal, and essentially none of this pre-processing makes sense, since they send a raw

byte-stream with no desire for line buffering or control-flow protocols. Thus, the biggest challenge in using the devices correctly was turning off all of the terminal-related overhead. Another key issue was correctly setting the speed and other communication parameters.

The principal input hardware was a head mounted Fastrak rotation/position sensor and a processing unit that interprets data from the sensor and sends it to the workstation via an RS-232 serial line. It was clear that the control process should listen to the serial port in raw mode, and as position/orientation data becomes available translate it into a position and orientation in model space, and finally send it to the simulation process as described above.

The secondary control interface is through a serial joystick. The joystick connects to a serial port and provides X-Y position information, and an additional coordinate derived from a knob on the side of the joystick. Several modes for use of the joystick are plausible interfaces for walking, and the decision on the actual walking interface was deferred until the implementation stage.

Communication with Real-time Rendering Software

The process that interfaces with the devices and calculates position data communicates with the rendering system through shared memory. This allows high bandwidth communication but allows the application to run as a separate process from the rendering engine to avoid interfering with the rendering pipeline. Very few problems resulted from using the shared memory interface, except on rare occasions when needed functionality could not be accessed through it.

Coordinate Transform of Input Device Data

In addition to interfacing with existing hardware and software, a number of other issues must be considered. First, coordinates from the head tracker must be translated into a coordinate system suitable for use with the simulation, and similarly for joystick coordinates. These must also be combined in a coherent way.

The position/rotation sensor sends its position and orientation relative to some initial state. However, these coordinates cannot be used directly to control the simulation. They must be mapped relative to the current camera position and orientation in the virtual space, and some accounting must be made for the fact that the sensor is not mounted between the eyes but rather on top of the head. This is simply a

matter of applying suitable transforms to the coordinates received. However, it is also necessary to initially calibrate the sensor. These issues also apply to the joystick except that the initial state is known ahead of time so calibration is unnecessary.

No single interface for combining the joystick and tracker coordinates met all of the projects needs. Thus, three different interfaces were implemented. The first is a "wheelchair" style arrangement with independent head motion. A current facing of the wheelchair is maintained internally by the program, and moving the joystick sideways rotates the wheelchair in place. Pushing the joystick forward or back moves the subject along the current wheelchair facing. The position and orientation of the viewer's head is taken relative to that of the wheelchair. Thus, it is possible to look sideways and move forwards, and see the scenery on one side pass by.

This first interface was considered potentially unsuitable for acquisition of spatial data. Thus, two more were implemented which did not maintain a separate orientation for the point of view and the direction of forward motion. The second interface completely ignores the sideways motion of the joystick, and allows forward or backward motion to move the subject along the current point of view (with the z coordinate removed to avoid up and down motion). The third is essentially the same except that moving the joystick sideways moves the subject sideways relative to the current point of view, allowing side-stepping.

Implementation of Collision Detection

During the walk-through, the user must not be permitted to pass through solid objects in the model. The available software already provides the means to do this collision detection, but it must be suitably enabled, and considered when controlling the motion of the point of view. Unfortunately, full collision-detection functionality is not available through the shared memory interface. Since the existing architecture was already working smoothly, a plan was devised that would allow it to work with only minimal modifications to the rendering engine. Essentially, a function in the simulation receives collision data, however, it communicates it through a separate IPC channel to the controlling process, which then handles collisions when they occur. The controlling process is able to directly enable collision detection and set collision geometries, but must resort to this back channel to be informed when collisions occur.

The virtual environment to be used includes stairs. This presents a special difficulty, since stairs must not be detected as a source of collision, but rather

should gradually elevate the position of the point of view, while allowing forward motion to continue. Since the simulation should work with minimal foreknowledge of the target environment, and since the simulation does not directly have data on all objects, solving this problem is difficult.

Support Efforts

Room-Scanning Robot

The room scanner (see figures 3 and 4) consists of a Quick-cam video camera connected to an Apple Macintosh computer and placed on a rolling cart with a stepper motor attached to its wheel. The computer controls the stepper motor, which advances the device allowing the Quick-cam camera to scan in a graphical texture image. The image and position data (derived from the stepper motor) is stored in a multitrack Quicktime file format as a sequence of 1-by-640-pixel frames. These single pixel-width images are then reconstructed into strips of texture for texture mapping on to the DWB model.

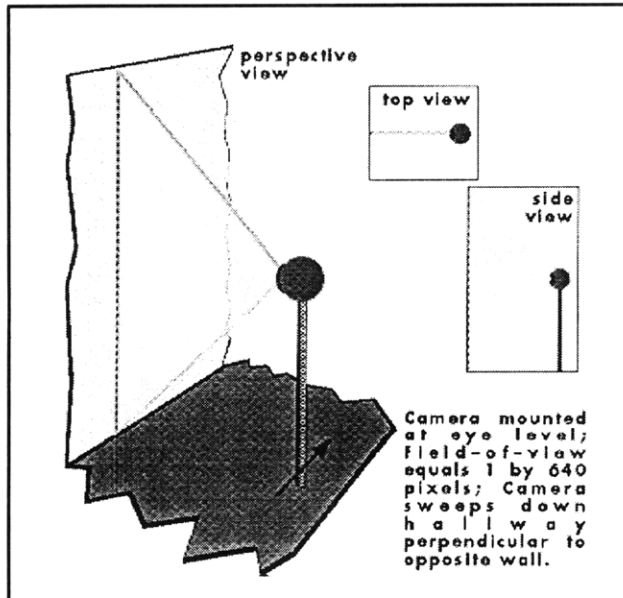


Figure 3. Room scanner conceptual plan.

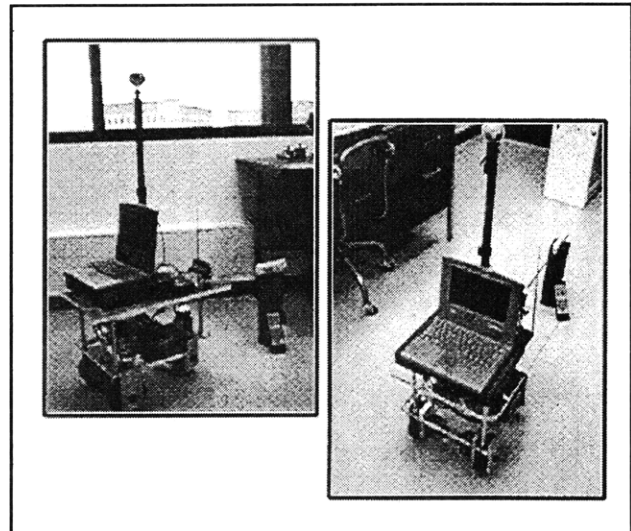


Figure 4. Room scanner prototype.

Mobility Interfaces

In addition to the room-scanning robot, we are exploring a number of "slippery-floor" mobility interfaces to enable the subject to move through the virtual environment by walking in place. One version of such an interface has been previously developed in Japan.³⁷ Prior to building our own version of such an interface, we are building a miniature version in which the subject's fingers do the "walking." We are interested in comparing the effectiveness of these two interfaces to each other as well as to other types of interfaces used to control motion through virtual space (control gloves, joysticks, etc.).

A Low-Cost Version of the Virtual Space

In addition to the high-fidelity virtual spaces, we are constructing a low-cost version that makes use of a readily available shareware-based development system and runs on a desktop computer. In this project, we are making use of Quake, the newest first-person perspective game from id.

³⁷ H. Iwata and T. Fujii, "Virtual Perambulator: a Novel Interface for Locomotion in Virtual Environment," *Proceedings of VRAIS'96*, 1996, pp. 60-65.

1.10 Auditory Cues to Combat Spatial Disorientation

Sponsor

U.S. Air Force - Office of Scientific Research
Grant F49620-95-1-0176³⁸

Project Staff

Nathaniel I. Durlach, Professor Richard M. Held,
Rabi Karmacharya

1.10.1 Overview

During the first year of this project (see *RLE Progress Report No. 138*), we completed development of the required facilities except for a controllable tilting bed that would enable the experimenter to have real-time control over subject orientation in the rotating room used to create acceleratory forces. The experiments performed during this first year established that perceived location of auditory images is affected in an orderly manner by the gravito-inertial vector. More specifically, when the vector simultaneously increases in magnitude above 1 G and rotates in the azimuthal plane of the head, the perceived azimuth of an auditory target moves in a direction opposite to that in which the vector moves.

Work during the second year has been directed towards a quantitative assessment of how the shift in perceived azimuth depends on the magnitude and angle of the gravito-inertial vector. In pursuit of this objective, we have completed design and implementation of the tilting bed and initiated a variety of new experiments. Results to date indicate that (1) a change in the magnitude of the vector with the angle fixed at 60 degrees has no effect, and (2) a change of head orientation in a 1 G field has only very minor effects. Experiments are now underway to determine the effects of a change in angle for a fixed vector magnitude in excess of 1 G, as well as to extend the results mentioned above to angles other than 60 degrees.

1.11 Research on Superauditory Localization for Improved Human-machine Interfaces

Sponsor

U.S. Air Force - Office of Scientific Research
Grant F49620-96-1-0202

Project Staff

Nathaniel I. Durlach, Professor Richard M. Held,
Dr. Barbara G. Shinn-Cunningham, Douglas S. Brungart, Gregory G. Lin, Kinuko Masaki, John Park, Salim F. Kassem, Elias A. Vyzas

1.11.1 Objectives

The general goals of this project are (1) to determine, understand, and model the perceptual effects of altered auditory localization cues, and (2) to design, construct, and evaluate cue alterations that can be used to improve performance of human-machine interfaces in virtual-environment and teleoperation systems. To the extent that the research is successful, it will both advance our understanding of auditory localization and adaptation, and improve our ability to design human-machine interfaces that provide a high level of performance. The specific goals for the project are listed below.

1. Continue to acquire, develop, and integrate devices and facilities into our laboratory that are suitable for studying supernormal auditory localization.
2. Analyze in more detail the data already obtained using the azimuthal remapping transformation

$$\theta' = f_n(\theta) = \frac{1}{2} \tan^{-1}$$

$$\left[\frac{2n \sin(2\theta)}{1 - n^2 + (1 + n^2) \cos(2\theta)} \right]$$

with $n = 3$.

3. Conduct additional experiments using the $f_3(\theta)$ transformation to clarify and broaden the results already obtained with this transformation, focusing primarily on how visual cues affect adaptation.

³⁸ Subcontract to MIT from Brandeis University, Waltham, Massachusetts.

4. Perform a series of experiments using the family of transformations $\{f_n(\theta)\}$ $n = 1, 2, 3, 4$, to study the effects of transformation severity and incremental exposure, as well as to explore questions related to conditional or dual adaptation.
5. Conduct a study similar to that mentioned in item 3 using the frequency-scaling family of transformations, which simulate increased head size, rather than the $f_n(\theta)$ family.
6. Refine our quantitative model showing how resolution and bias in the localization of azimuth are influenced by transformations of azimuthal localization cues.
7. Conduct a further series of experiments using the frequency-scaling transformations to determine the effects of such transformations on the perception of elevation.
8. Determine the potential of varying sets of filter transfer characteristics for coding distance and/or elevation by studying the information transfer that can be achieved using these sets.
9. Determine how the transformations that appear most promising in the previous studies, which evaluate the transformation in acoustic environments containing only a single target, perform in multiple-simultaneous-target environments.
10. Using one of the more promising sets of filter transfer characteristics studied in project 8, evaluate the usefulness of this set for coding distance and/or elevation by conducting adaptation experiments similar to those pursued in projects 3, 4, and 5 on the azimuthal variable.
11. Extend our knowledge of natural localization cues by measuring and analyzing head-related transfer functions (HRTFs) for sources in the near field (i.e., at a distance of 1 meter or less from the center of the head).

1.11.2 Progress Report

During the past year, progress has been towards goals 1, 3, 4, 6, and 11. In addition, we have completed three manuscripts describing work completed under a previous U.S. Air Force, Office of Sponsored Research grant.

Our model of adaptation to altered auditory localiza-

tion cues has been modified, and the resulting model has been tested with all previous experimental results (item 6). The modification caused little change in predicted results (both the original model and the refined model fit resolution and bias results quite well); however, conceptually, the model is simpler and more manageable than the original model. In addition, we have completed experiments investigating the role of visual cues on auditory adaptation (item 4) and the effect of changing the auditory transformation during an experimental session (item 3). Development of appropriate software to allow us to use the Tucker-Davis Power-DAC in future experiments has progressed (item 1), although some effort remains before these experiments can proceed. Finally, we have begun to measure (and to test the use of) HRTFs for sources located close to the head (i.e., at a distance of 1 meter or less) and have performed pilot experiments to evaluate different response methods for performing near-field (distance) localization measurements (item 11).

Accomplishments/New Findings

Model Predictions

A detailed description of our model of adaptation to supernormal auditory localization cues can be found in a paper being prepared.³⁹ The main features of the model are summarized below.

The model assumes that all responses are determined by the value of an internal decision variable which is stochastic. The mean of the decision variable is monotonically related to the azimuth normally associated with a given stimulus, while its variance depends upon the range of stimuli being attended by the subject at a given point in time. More specifically, the variance in the underlying decision space grows as the attended range of stimuli grows. Criteria are placed along the uni-dimensional decision axis to divide the axis into n contiguous regions, corresponding to the n possible responses on the n -alternative, forced-choice identification tasks used in our study. Adaptation occurs as the criteria shift, allowing subjects to change their mean response to specific stimuli as the locations of the response regions move.

The model assumes that placement of the decision criteria and effective range (determining underlying variability in the decision variable) are determined by the relationship between physical stimuli and

³⁹ B.G. Shinn-Cunningham, "Adaptation to Supernormal Auditory Localization Cues: A Decision-theory Model," *Percept. Psychophys.*, in preparation.

mean response at any given point in time. These changes are examined fully in a paper that shows that mean response is linearly related to the azimuth normally associated with a particular physical stimulus at all times.⁴⁰ Changes in mean response occur as the slope relating these values exponentially approaches an asymptotic value. The rate of change (reflecting the rate of adaptation) is statistically independent of the experimental factors investigated to date, while the asymptote depends upon the strength of the transformation employed.

With these assumptions, our model is able to fit total sensitivity Δ' (sum of δ'_i across all adjacent pairs of stimuli in an experiment), bias, and resolution for all experiments performed to date. The three figures below show the effectiveness of the model in fitting these quantities. It should be pointed out that the model parameters were chosen to fit Δ' (total sensitivity). The resulting parameter values were then used to predict bias and resolution (with good results).

Figure 5 shows the actual results (error bars) and the model predictions (open circles) for total sensitivity as a function of run. For each experiment, the two free model parameters were fit by finding values which minimized the mean square error between predictions and total sensitivity for each subject. These parameter values were then averaged across subjects to yield the model parameters used in all predictions. The figure shows predictions and results for five experiments, which differed mainly in the strength of the transformation employed ($n=2, 3$, or 4) and the number of source positions used (either $[-60, +60]$ degs or $[-30, 30]$ degs). Details of the differences across experiments can be found in a submitted paper.⁴¹

Predictions match overall magnitude of the results well, although there is a tendency to underestimate sensitivity in some experiments (experiments III and V, panels a and c) and to overestimate sensitivity in experiment VI (panel d). Identical model parameters were used to fit data from all experiments; thus, the fit across all experiments should be con-

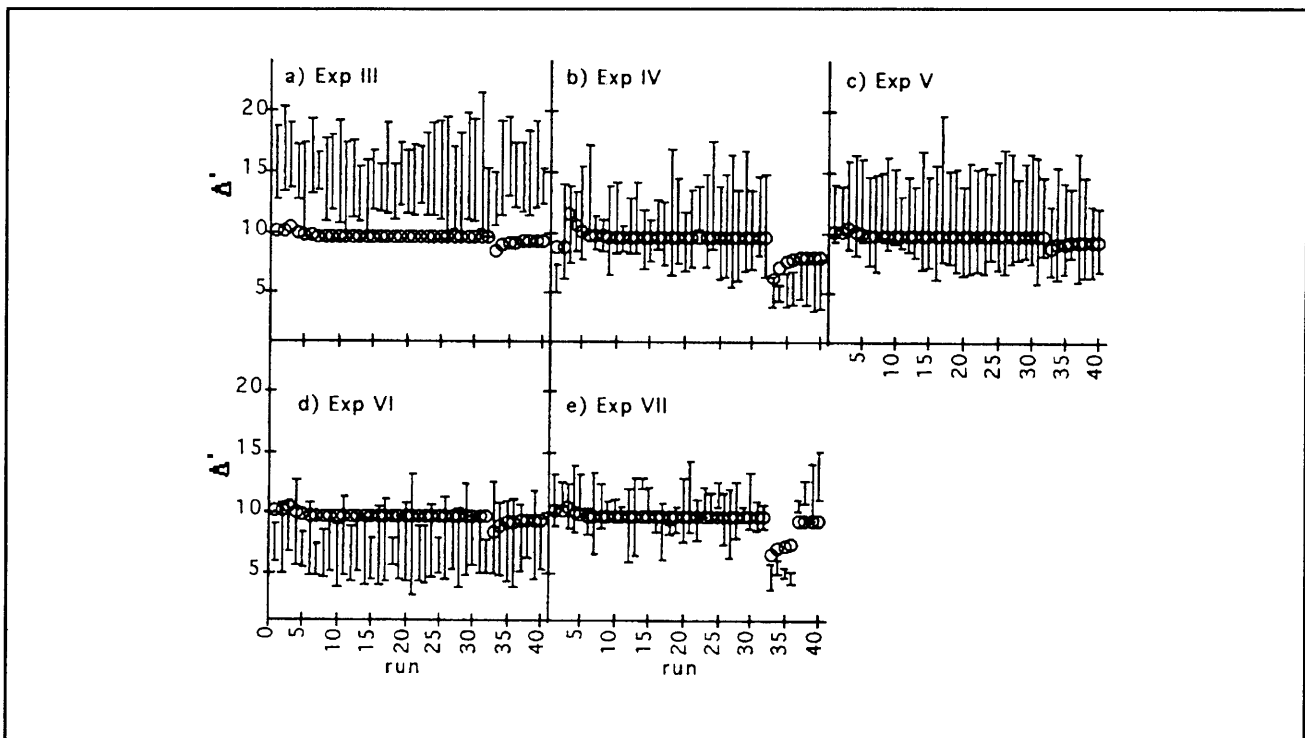


Figure 5. Predicted and actual Δ' as a function of run (experimental data repeated from figure 4). Open circles show predicted values, solid lines give mean experimental results (averaged across subjects) plus and minus one standard deviation.

⁴⁰ B.G. Shinn-Cunningham, N.I. Durlach, and R.H. Held, "Adaptation to Supernormal Auditory Localization Cues II: Changes in Mean Response," submitted to *J. Acoust. Soc. Am.*.

⁴¹ B.G. Shinn-Cunningham, N.I. Durlach, and R.M. Held, "Adaptation to Supernormal Auditory Localization Cues I: Bias and resolution," submitted to *J. Acoust. Soc. Am.*

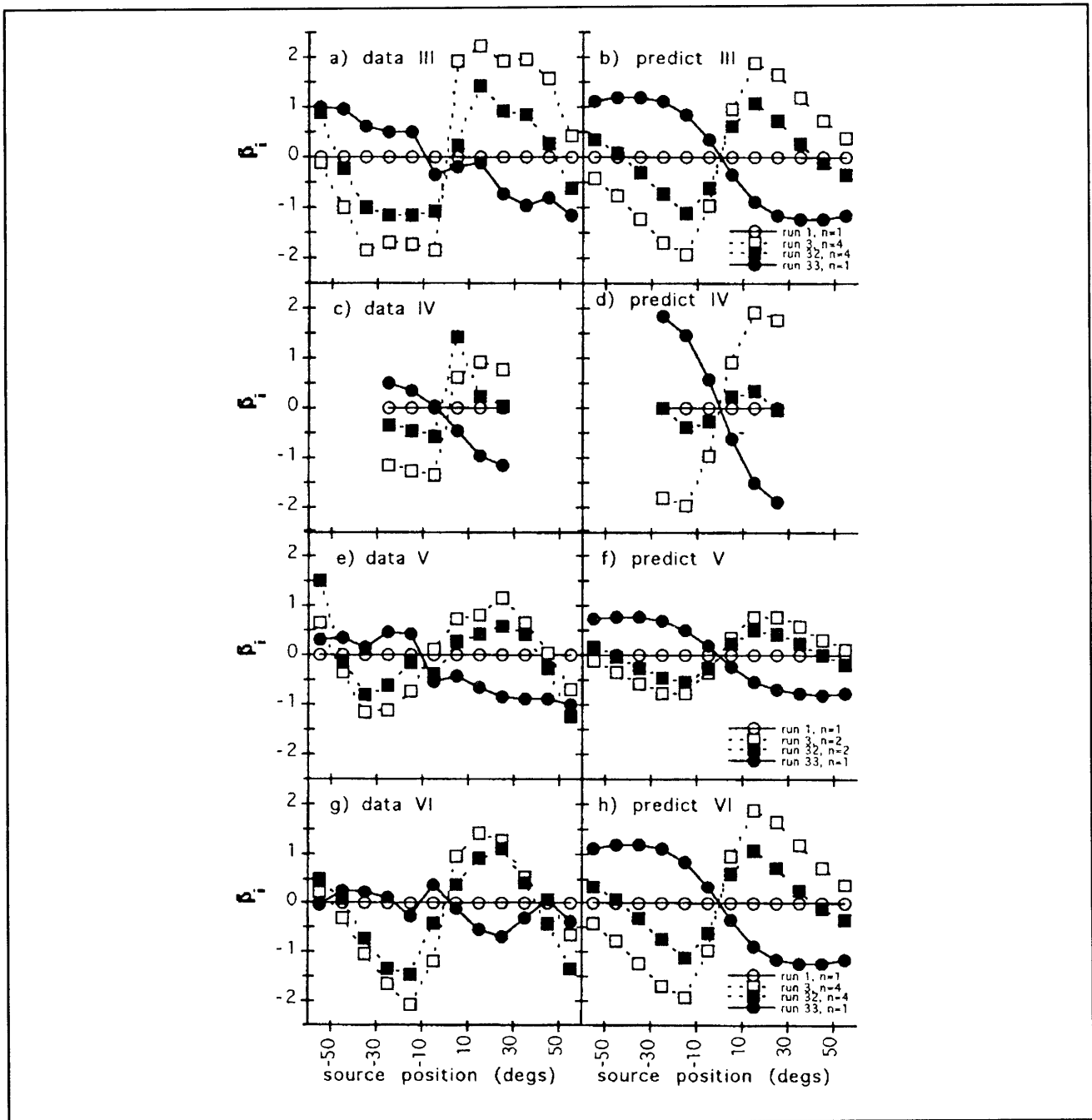


Figure 6. Actual and predicted bias as a function of position. Left panels show experimental results, right panels show corresponding model predictions. Open symbols represent results prior to training; filled symbols represent results after supernormal exposure. Circles represent normal-cue results; squares show altered-cue results.

sidered. The model also predicts abrupt changes in total sensitivity quite well (note results and predictions for Experiments IV and VII, panels b and e).

Figure 6 shows the actual and predicted results for bias β_i as a function of source position for the same experiments. In this figure and in figure 7, open symbols show data prior to adaptation and filled

symbols show results after adaptation. Circles represent normal-cue results and squares represent altered-cue results. In figure 6, actual results are shown in the left panels and the corresponding model predictions shown in the right panels. The degree to which the left and right panels are alike is a measure of the degree to which the model predicts bias correctly. In all experiments, a large bias is introduced when altered cues are first presented

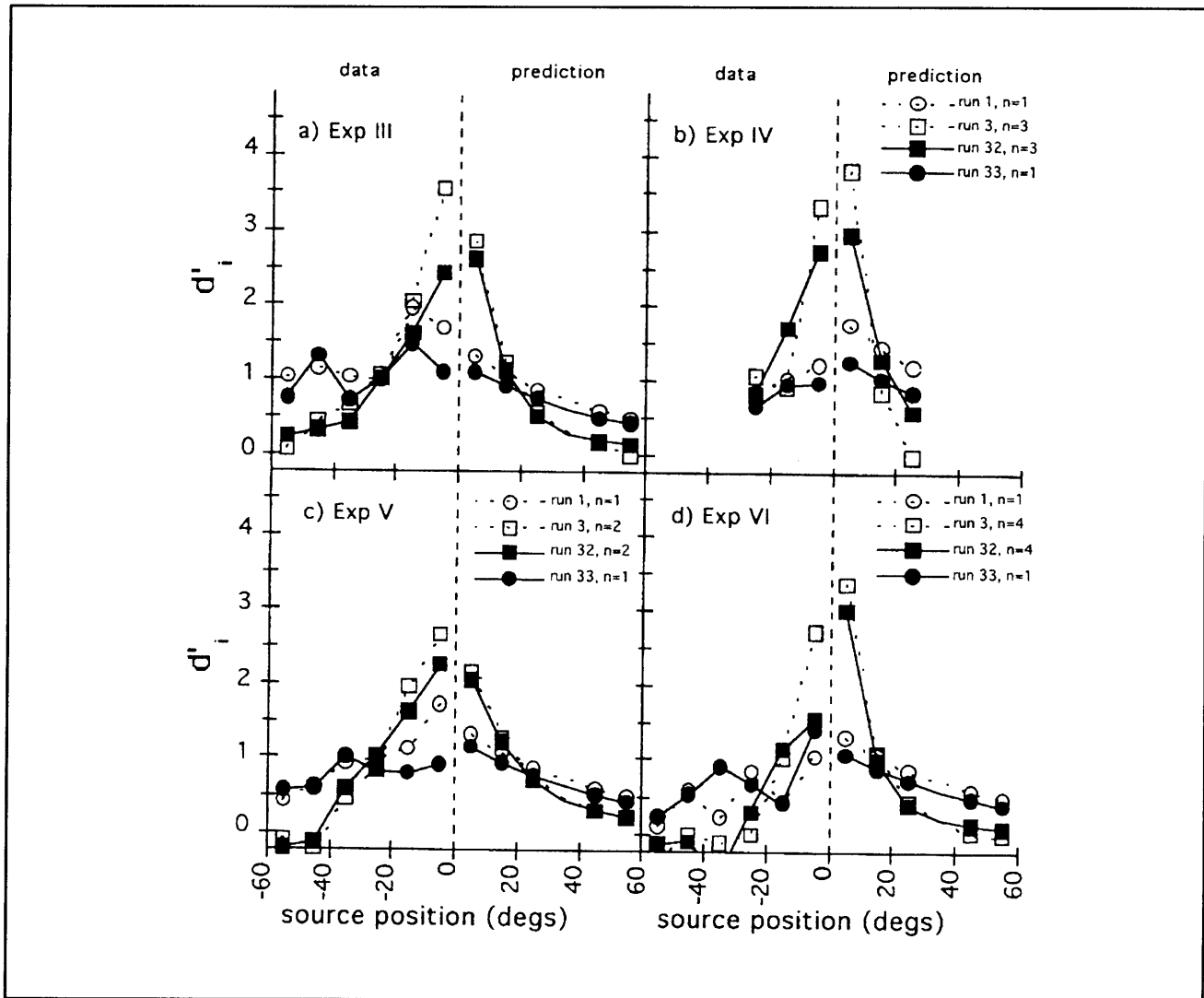


Figure 7. Actual and predicted δ'_i as a function of position. Left side of each panel shows experimental results (averaged assuming data are left-right mirror symmetric). Right side of each panel gives predictions from the model for positions right of center. The dotted line separates the experimental data from the predicted results. Closeness of predictions is measured by the degree to which each panel is mirror symmetric. Again, the model predicts results quite closely. Model parameters were identical to the values used to predict total sensitivity (figure 5) and bias (figure 6).

(open squares). This bias is reduced as subjects adapt (compare filled to open squares). At the end of the adaptation period, a large bias is seen which is opposite in sign to the bias first introduced with the altered cues (compare filled circles to open squares). Model predictions show the same trends for all experiments.

Finally, figure 7 shows results and predictions for resolution δ'_i as a function of source position. In figure 7, the left side of each panel shows experimental results (averaged assuming data are left-right mirror symmetric) and the right side of each panel gives predictions from the model for positions right of center. The ability of the model to predict

these results is measured by the degree to which each panel is mirror symmetric. Model parameters were identical to the values used to predict total sensitivity (figure 5) and bias (figure 6).

Again, the model predicts results quite closely. The use of supernormal cues produces an increase in resolution for positions near zero degrees azimuth (open squares). As subjects adapt to the supernormal cues, however, there is a slight tendency for resolution to decrease (compare filled to open squares). After adaptation, resolution using normal localization cues also tends to be worse than resolution prior to training (compare filled to open circles). The model predicts all these trends. In

addition, quantitative differences between experiments are also predicted by the model.⁴²

Effect of Visual Cues

Two previous experiments showed that adaptation to auditory-cue transformations depends upon visual information. In one experiment, subjects were blindfolded throughout the experimental session. When blindfolded, subjects could theoretically obtain information about the auditory-cue transformation by comparing the felt position of the head with auditory localization information. In the control experiment, subjects had a full visual field available to them (which could improve the accuracy with which the head position was registered) and were provided with explicit visual information (via a small lightbulb) as to the azimuthal position of the heard auditory source. In the blindfolded experiment, subjects showed no adaptation: localization errors were unchanged after 40 minutes of exposure to supernormal cues. In the control experiment, adaptation was found in all subjects. Two additional experiments have been performed to test exactly what type of visual information is necessary to allow subjects to adapt to transformed auditory localization cues.

The first experiment was identical to the previous blindfolded experiment, except that subjects viewed a normal visual field throughout the experiment, and the possible locations of the auditory source were marked by 13 lights at the 13 possible source positions. Unlike the previous control experiment, subjects were not given any explicit visual information about the actual location of the heard auditory source (since the lights were never turned on). Preliminary analysis shows that subjects in this experiment adapted as fully as did subjects who were presented with explicit visual information about the location of the heard auditory source.

In the second experiment, the light arc (denoting the possible locations of the auditory sources) was removed. Thus, in these experiments, subjects had a complex visual field available to them at all times, but had no visual information about either the correct auditory source location, or the possible locations of the auditory sources. Preliminary analysis of these results indicate that subjects adapted in this condition, but that the adaptation was less complete than in the condition when the light arc marked all possible source locations.

Further analysis is necessary to fully quantify differences among these experiments; however, taken as a whole, these results indicate that the felt position of the head is not a sufficient cue to allow subjects to learn auditory cue rearrangements (cf., results from experiment with blindfolded individuals). However, when a visual field is present, subjects may be able to accurately register the location of their heads and can thus deduce how auditory localization cues change with changes in head position (cf., results from two new experiments). Finally, visual information about the possible locations of auditory sources also helps subjects to learn auditory cue remappings (compare results of two new experiments). Any additional benefit provided by explicit visual information as to the correct location of the auditory source is too small to be measured.

Effect of Changing Auditory-Cue Transformation

Two additional experiments have been completed which investigate what occurs when the strength of the transformation is changed half-way through the training period. In one experiment, a transformation of $n=2$ was used for the first 15 altered-cue tests and a transformation of $n=4$ was used for the final 15 altered-cue tests. In the second experiment, the order of the two transformations was reversed, with an $n=4$ transformation presented first and an $n=2$ transformation presented at the end of the training period.

Preliminary analysis suggests that, similar to results for single-transformation experiments, adaptation to multiple transformations can be measured by measuring the slope relating mean perceived location and physical cue. As in the single-transformation experiments, the slope appears to change exponentially towards an asymptotic value that depends only on the strength of the current transformation. Thus, in the $n=2$, $n=4$ experiment, subjects show an exponential decrease in slope towards the asymptote for $n=2$, then show a further exponential decrease in slope towards the asymptote for the transformation $n=4$. In the $n=4$, $n=2$ experiment, subjects show an exponential decrease towards the asymptote for $n=4$, then show an exponential increase in slope towards the asymptote for $n=2$. These changes in mean response can be fit by simple extension of the exponential curve-fitting

⁴² B.G. Shinn-Cunningham, "Adaptation to Supernormal Auditory Localization Cues: A Decision-theory Model," *Percept. Psychophys.*, in preparation.

performed for single-transformation experiments.⁴³ Further analysis is necessary to test whether bias and resolution results from these experiments can be predicted by our model of adaptation.

Near-Field Head-Related Transfer Functions

In order to extend our knowledge of natural localization cues, we have begun to measure (and to test the use of) head-related transfer functions (HRTFs) for sources in the near field (i.e., at a distance of 1 meter or less). The localization cues that are available in this "near-field" region are different than those in the far field and have never been carefully studied. The results of this work are expected not only to increase our knowledge about the perception of direction for sources in the near field, but also to provide us with important information concerning the perception of distance (both for natural cue situations and for the development of supernormal distance cues).

Our study of near-field auditory localization cues has progressed in two major areas. A KEMAR acoustic manikin has been used to collect HRTFs for distances ranging from 0.15 m to 1 m from the center of the head, in cooperation with the Bioacoustics and Biocommunication branch of the Crew Systems Directorate of the Armstrong Laboratory (AL/CFBA) at Wright Patterson Air Force Base in Springfield, Ohio. Measurements include azimuths and elevations from -45 degrees to 45 degrees. The most striking difference between these HRTFs and far-field HRTFs is an increase in interaural intensity differences (IIDs) as the sound source approaches the head. Further work will investigate whether these distance-dependent IIDs may allow listeners to determine source distance in the near field without reliance on overall intensity cues.

A pilot experiment designed to evaluate possible response methods for near-field auditory localization experiments was completed. Four response methods were compared: (1) reporting coordinates verbally, (2) pointing directly to the perceived location with a sensor on a wand, (3) pointing to the perceived location relative to the location of a full-sized manikin head, and (4) pointing to the perceived location relative to the location of a half-sized manikin head. Initial data analysis indicates that performance is best in the simple pointing task and worst when subjects must point to the location relative to the full-size manikin head.

Summary

Substantial progress has been made in a number of areas. Our preliminary model has been refined and tested against all previous experimental results with good success. Three manuscripts describing our experimental results, analysis, and model have been prepared for submission (see publications list). Four additional experiments have been completed investigating two different areas of substantial theoretical importance for the project. Preliminary analysis suggests that the current model of auditory adaptation can be extended to predict the results of these new experiments. Finally, acoustic measurements have been collected and preliminary experiments completed to support our examination of auditory localization cues in the near field.

1.11.3 Publications

Shinn-Cunningham, B.G., N.I. Durlach, and R.M. Held. "Adaptation to Supernormal Auditory Localization Cues I: Bias and resolution." Submitted to *J. Acoust. Soc. Am.*

Shinn-Cunningham, B.G., Durlach, N.I., and R.M. Held, "Adaptation to Supernormal Auditory Localization Cues II: Changes in Mean Response." submitted to *J. Acoust. Soc. Am.*

Shinn-Cunningham, B.G. "Adaptation to Supernormal Auditory Localization Cues: A Decision-theory Model." *Percept. Psychophys.* In preparation.

⁴³ B.G. Shinn-Cunningham, N.I. Durlach, and R.H. Held, "Adaptation to Supernormal Auditory Localization Cues II: Changes in Mean Response," submitted to *J. Acoust. Soc. Am.*

1.12 Training for Remote Sensing and Manipulation

Sponsor

U.S. Navy - Office of Naval Research
Subcontract 40167⁴⁴

Project Staff

Nathaniel I. Durlach, Dr. David Zeltzer, Dr. Thomas E.v. Wiegand, Walter A. Aviles, Dorrie Hall, Francis G. Taylor, Jonathan D. Pfautz, David H. Manowitz, Rebecca L. Garnett

1.12.1 Summary

The Training for Remote Sensing and Manipulation (TRANSoM) program is an interdisciplinary research effort to design, develop, and evaluate the use of virtual environment (VE) systems for training and mission rehearsal for teleoperated mini-submarines, often called remotely operated vehicles (ROVs). Intelligent tutoring system (ITS) techniques will be incorporated in the ROV pilot training and mission rehearsal VE system, which will be demonstrated in the operation of single and multiple ROVs in a shallow water mine countermeasures (MCM) mission. The variety of skills demanded for ROV operation, including piloting and navigation, close-in maneuvering, manipulation, supervisory control, situation assessment, and team interaction, provides a unique opportunity for combining the strengths of ITS and VE technologies. A three to five year effort is planned in VE-based ITS research and development.

Research efforts are concentrated in four main areas: (1) understanding the cognitive and sensorimotor skills required for the successful utilization of remote sensor and manipulator platforms, (2) extending ITS theory and techniques to encompass sensorimotor as well as cognitive tasks, and developing an ITS part-task training tool kit, (3) exploring Human-Machine interaction (HMI) tools, techniques and modalities for their effects on training, and (4) validating training effectiveness and transfer using both VE-based simulations and actual ROV systems in part-task and full mission scenarios.

Technological efforts to support the research consist of the following:

- Providing an ROV system to provide experimenters access to essential ROV controls and

displays to support operational testing in both tank and shallow water environments.

- Configuring existing computational and HMI systems to support training system development and experimentation. These include the facilities established for the U.S. Navy Virtual Environment Technology for Training (VETT) program.
- Providing a system to permit both remote ROV operation and tele-conferencing between the geographically distributed research and development team.

Three main areas of system transition are planned: (1) production of a prototype MCM ROV pilot trainer, (2) application of techniques and technologies to other military requirements such as surveillance, security, and autonomous underwater vehicles, and (3) development of systems for commercial ROV operations and training.

Five organizations are collaborating on this effort. Imetrix, Inc., is the prime contractor, overseeing the effort, providing the ROV and control expertise, and the main conduit for product transition. RLE's Sensory Communication group is providing the lead in human-machine interaction research and virtual environment technology. McDonnell Douglas Training Systems, St. Louis, Missouri, is providing expertise in the design of intelligent tutoring systems and leading the training transfer studies. BBN Corporation, Cambridge, Massachusetts, is developing the ITS and providing training research and evaluation support. Learning Research and Development Center, Pittsburgh, Pennsylvania, is providing training expertise in team interaction.

1.12.2 TRANSoM VE System Architecture and Functionality

In section 1.12.2 of this report, the required functionality and the concomitant software architecture of the TRANSoM VE system will be discussed. This will include a presentation of the various human-machine interfaces and computational models and processes that have been developed. In section 1.12.3, we will describe some associated experiments on the role of sensorimotor involvement in training transfer. Finally, in section 1.12.4, the management of the collaborative software development effort will be described, including implementation of software version control and World-Wide-Web-based development tools.

⁴⁴ Imetrix Corporation, Cetaumet, Massachusetts.

System Design

The term "baseline simulation" refers to the object space, viewer-independent definitions of component objects, agents and processes of a particular virtual world. Geometry and kinematics must be specified; descriptions of dynamic properties (e.g., mass, inertia, Young's modulus) are required for physically based models. In general, a baseline simulation may incorporate a variety of simplifications and algorithmic shortcuts needed to maintain interactive update rates, but the intent of the notion is to provide Q so far as possible with a given suite of hardware, and a desired level of representational complexity Q a realistic rendition of some physical environment.

It is through the human/machine interface to a baseline simulation that one directs the viewpoint or selects and manipulates objects. It is of course extremely difficult if not impossible to characterize all the possible enhancements to a baseline simulation. But by enhanced simulation we intend to account for such notions as the scaling of physical parameters, i.e., the gravitational constant; various perceptual distortions, such as computing visual imagery in stereo with an enlarged interocular distance to enhance the sense of depth; or various instructional interventions. Such enhancements are features that obviously distinguish VE training systems from conventional trainers based on physical mockups or out-of-service operational systems. With the inclusion of instructional features, the baseline ROV simulation becomes an enhanced simulation.

The primary tasks of Year 1 have focused on the development of the baseline ROV simulation, and a suite of rapid prototyping tools to enable the training specialists in the TRANSoM program to investigate the requirements and functionality of the TRANSoM ITS and to develop and evaluate candidate-instructional interventions. The primary user of the TRANSoM VE system in Year 1, therefore, has been the developer the VE specialists responsible for designing, implementing and evaluating the TRANSoM VE system itself and the training specialists responsible for the ITS. These efforts have involved requirements engineering to determine the nature of the virtual worlds to be implemented, and the desired functionality of the trainees', developers', and instructors' interfaces; development of hardware and software specifications derived from the requirements; and system design, implementation and evaluation. This work has been completed.

TRANSoM Modules

The TRANSoM system consists of three main elements, (1) application models and processes, (2) logical interface. and (3) physical interface (figure 8).

As shown in figure 8, the physical interface consists of the sensors used to transmit the intentions of the developer to the VE, and the audio and visual displays by which information is presented to the developer. The logical interface represents the human-machine interface software that mediates the interactions among the major components of the system, manages the interface by which the developer communicates with, and receives information from the TRANSoM VE. The Computational models and processes required by the TRANSoM VE are enumerated and described in the Activity charts, detailed below. The VETTnet, speech and audio server, audio spatializer, and BBN hark voice input modules were developed as part of the VETT core testbed, and are discussed in detail in the VETT core testbed Year 1 technical report. The interested reader is referred to that publication.

Activity Charts

Activity charts show the functional architecture and relationships among the modules of the TRANSoM VE system. At the highest level, the TRANSoM activities chart looks much like the TRANSoM Module chart (see figure 9). The design of the TRANSoM VE is based on extensive working sessions with members of the TRANSoM team, including researchers from Imetrix, McDonnell Douglas Training Systems, and BBN. These deliberations included discussions of experiences gained from visits to an operational U.S. Navy ROV facility in San Diego, and a U.S. Navy ROV training site in Florida, as well as the operational experience of the Imetrix investigators. In particular, researchers from MIT and BBN held weekly meetings, often joined by engineers from Imetrix. The focus of these meetings was to design, implement and evaluate the TRANSoM system, especially the rapid prototyping developers' interface, the ROV operator interface, training mission scenarios and the concomitant VE system requirements, and candidate instructional interventions. Figure 10 shows a typical view of a TRANSoM training session, and this view includes elements from all the modules described.

As shown in figure 11, the TRANSoM virtual world is composed of the Undersea Environment, the ROV, which generates various ROV paths, and the ROV console by which the operator Q developer or trainee Q controls the ROV.

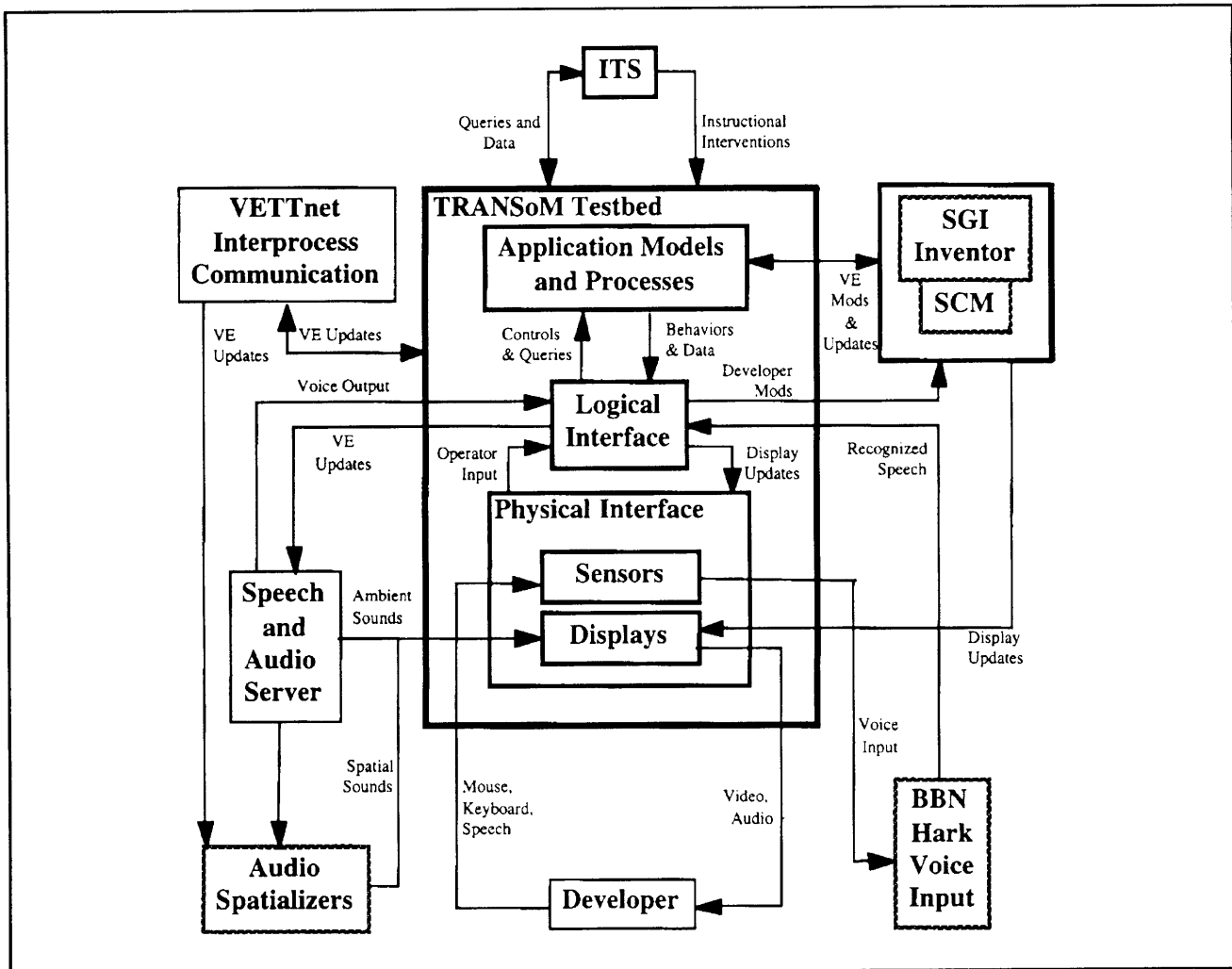


Figure 8. Major TRANSOM modules.

Interaction: Logical Interface

The Logical Interface in the current version of TRANSOM is managed by the Developers' Interface. During Year 1, a number of interpretive languages Q Python, JAVA, GUILLE, ALICE, were evaluated as front-ends for the SGI Inventor software which controls graphical functions on the SGI Indigo platforms being used for TRANSOM system development. In the end, the Scheme language Q, a statically-scoped version of Lisp Q was chosen as the interpretive front-end. The resulting system provides nearly complete access to the realtime graphics controlled by Inventor, and at the same time, allows the virtual world to be easily defined and modified at run-time if desired.

The Ivy-Scheme system, developed by MIT Professor Ken Russel, was eventually selected as the interpretive front-end. A number of significant changes, however, were made to Ivy-Scheme in order to adapt it for use in the TRANSOM program.

The header2scheme utility, also developed by Russel, is used to give Ivy-Scheme transparent access to C++ objects and procedures, and thus allows SGI Open Inventor to be accessed at run time through Ivy-Scheme. The X window system is also callable through Ivy-Scheme, which makes available the Motif library for rapid interface prototyping.

Direct manipulation is not always the most productive style of interaction, especially when complex virtual worlds are involved. In particular, unconstrained, interactive navigation through large virtual worlds can be difficult and disorienting. It is especially crucial in a training application not to burden the trainee with the additional task of controlling the viewpoint.

Views are carefully managed in the prototype trainees' interface. A view from the simulated ROV video camera is always shown in the upper left. The Developers' Interface allows the visibility atten-

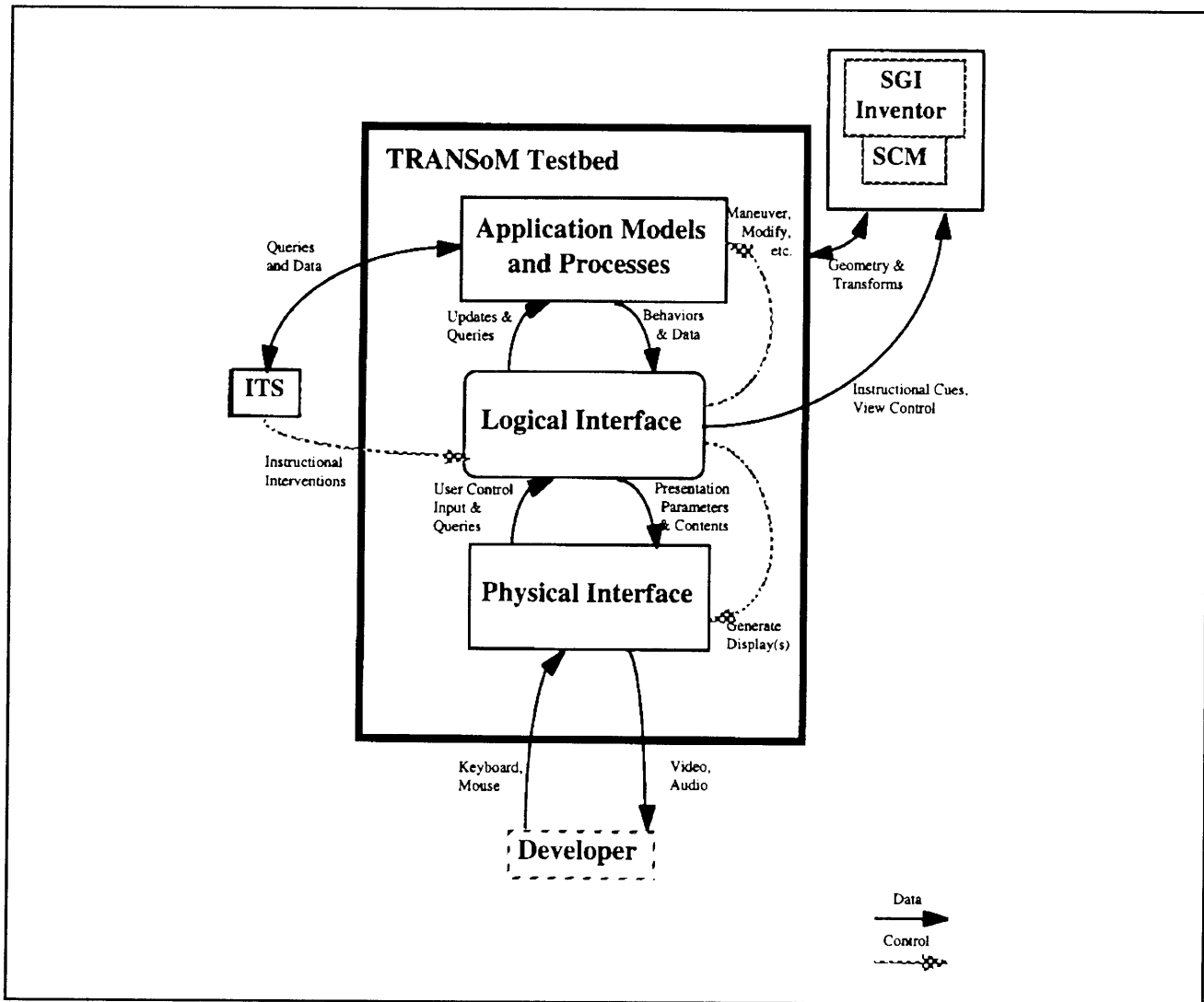


Figure 9. TRANSoM activities.

uation and the amount of suspended particulates to be easily varied for instructional purposes. The upper right view always shows the Nav Plot, which indicates the ROV as a triangle and shows a snail trail of the ROVs maneuvers. The middle right view always displays vehicle status instruments, in particular, compass heading is shown with a compass rose, and the current vehicle depth, and video camera tilt and zoom are displayed as well. The GUI console is shown in the lower right, but the GUI console will be replaced early in Year 2 with the physical control console supplied by Imetrix.

In the lower left window, one of a three VE views is always displayed for instructional purposes. Currently, at system startup, a bird's eye view is displayed, which is simply a plan view of the undersea world as seen from a viewpoint centered directly overhead. This view provides overall visual context, but since detail are difficult to see, it is

usually replaced by either the Wingman or Stadium view.

The apparent visibility through the ROV video camera can be controlled through the Developers' Interface, either by altering the visibility attenuation factor Q which renders the undersea view increasingly "hazy" with distance, or by turning on or off suspended particulate. Underwater currents can also be toggled on or off, and combined with suspended particulate that move with the resultant of current motion and ROV motion, this provides strong directional cues.

A record and playback capability has also been implemented. In the current system, a time-history of all ROV maneuvers Q including pitch and zoom of the view camera Q are saved. At any point during a mission, but usually at the end of a mission, the mission, or any portion of it, can be

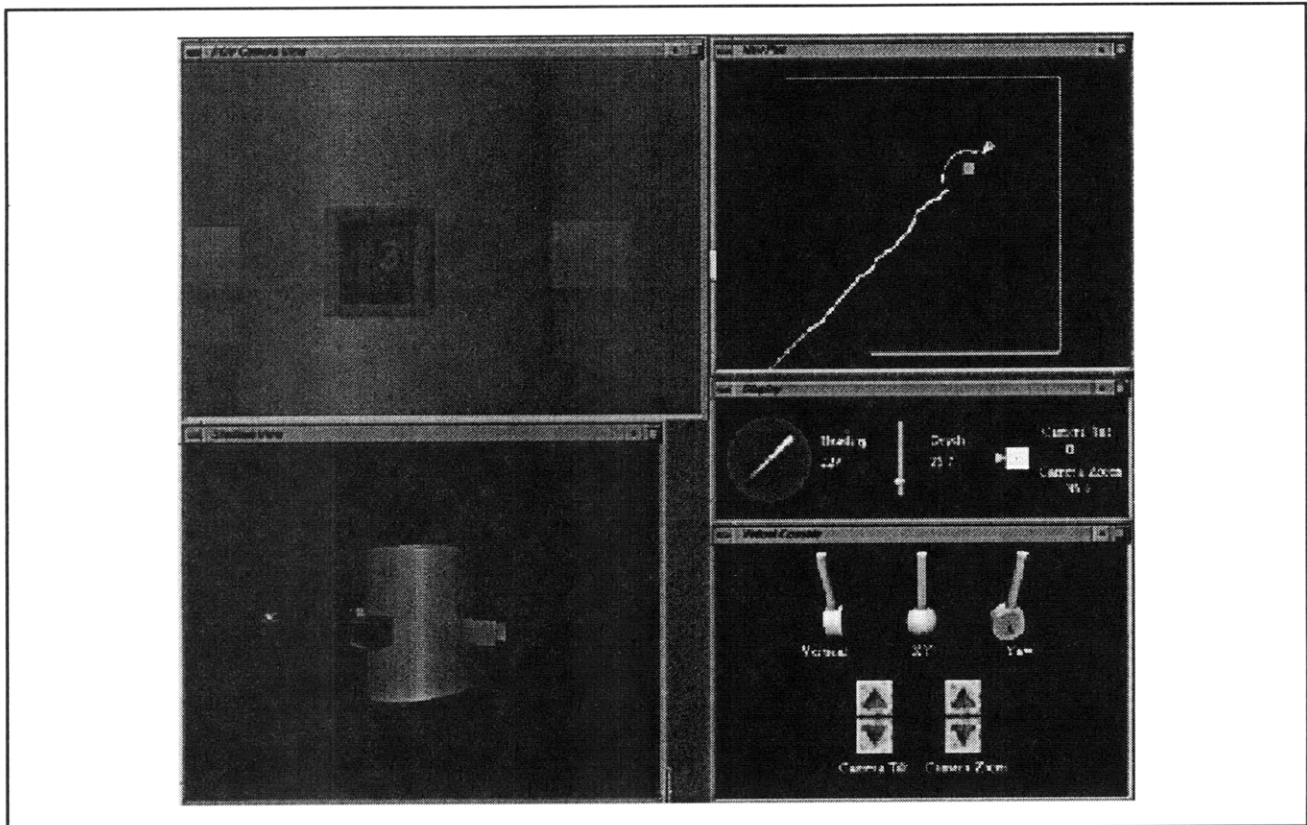


Figure 10. Typical view of a TRANSOM training mission showing the ROV near the target object. Imagery, as seen from the ROV video camera, is seen at the top left. The Stadium View Q, one of several specially constrained VE instructional views of the Undersea Environment Q, can be seen at the lower left. The straight line emanating from the ROV is the "camera laser," an instructional cue which indicates the direction of view of the ROV video camera. The Nav Plot and the Vehicle Status Display are shown in the upper right and middle right, respectively. The ROV is controlled using the virtual ROV Console, shown at the lower right. The console graphical user interface (GUI) will be replaced by the physical Imetrix console. These views are discussed further in the text.

replayed. This means that the ROV video camera view, the Nav Plot, and the GUI interface are re-displayed in realtime exactly as they appeared during the mission. Since the position and orientation of the vehicle are available, however, the VE views can be freely interchanged by the developer or the ITS from among the Bird's Eye, the Wingman or the Stadium views. "VCR" controls are provided to rewind, fast forward, or pause during playback.

Physical Interface

It is thought that a Head-Mounted Display (HMD) will be useful in certain part-task trainers. For this reason, an asynchronous server program has been designed, implemented, debugged, and integrated into the TRANSOM system for managing high-speed data input from RS-232 devices. While this is a general interface, it is designed to work efficiently with the Polhemus position tracker and the forthcoming Imetrix ROV joystick controller. This also

enables head-tracking for use with an HMD. Spoken output for coaching has been implemented and integrated into the current ITS. To provide a sound spatialization effect in the TRANSOM project, the following issues have been addressed:

1. Determining the number of sound sources to be used.
2. Choosing the right audio files for the sound sources.
3. Synchronizing the ROV event with the sound event.
4. Spatializing the virtual thruster sound.
5. Tuning the loudness of sound source according to the speed of thruster.
6. Changing the pitch of the sound source according to the acceleration of the thruster.

1.12.3 Sensorimotor Involvement Experiments

A fundamental focus of our research is the exploration of the effects of sensorimotor involvement on training transfer. As opposed to classical simulation-based training systems, VE-based training systems allow reconfiguration and manipulation of the "near-field" in the haptic as well as the visual and auditory modalities. In addition, simulated environments can be explored in such a way that movement through these spaces may be controlled with a variety of methods involving sensorimotor interaction (including "natural" methods such as walking or running). Therefore, the potential for a multi-modal, reconfigurable, near-field, and, to a lesser extent, sensorimotor involvement for movement and exploration of the simulated environment distinguishes VE-based training systems from previous training systems and technologies. The desired

form and utility of these various forms of sensorimotor involvement, however, are largely unexplored. Designing, building, and incorporating technologies which allow sensorimotor involvement (e.g., force-reflecting devices, locomotion platforms, etc.) severely impacts the cost and complexity of a VE-based training system. From a training perspective, these costs must be balanced against the demonstrated utility of the interventions for training transfer. The goal of our efforts is to explore the utility of sensorimotor involvement on two key skills which must be learned by a remotely operated vehicle (ROV) pilot. These skills are: (1) dynamic position and orientation control, and (2) path awareness.

In order to effectively control the position and orientation of an ROV, an operator must: (1) be familiar with the layout, information, and intended effects of the ROV's displays and controls, and (2) explicitly

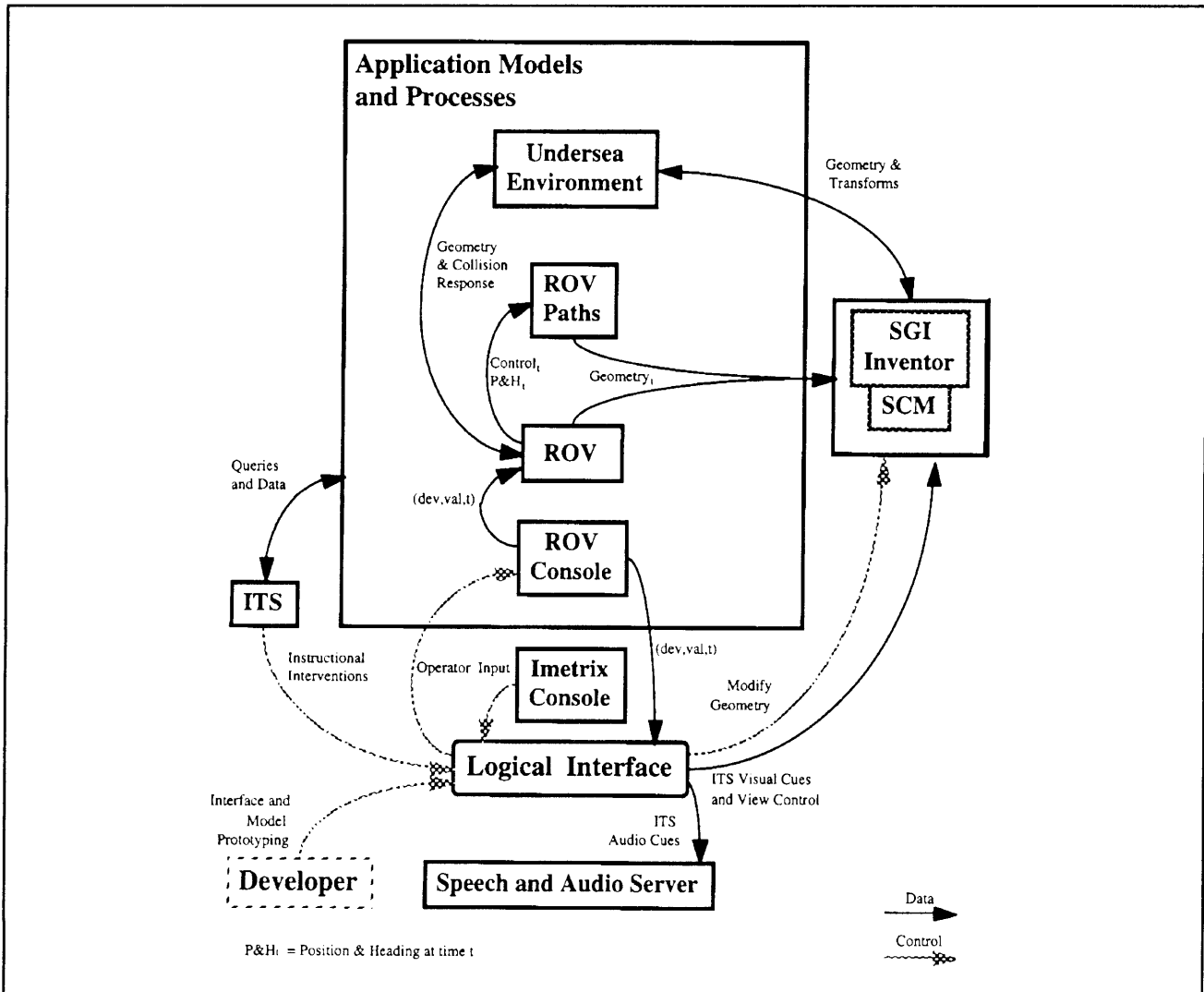


Figure 11. TRANSoM model activities.

or implicitly understand the ROV's response to control inputs over time (i.e., the dynamics of the ROV). Our research to date focuses on the latter. Specifically, we are exploring the effects of force-feedback on learning effective ROV dynamic position and orientation control. This research is motivated both by the importance of dynamic position/orientation control as a component of ROV piloting and by the fact that force-feedback devices capable of supporting this work are readily available. Furthermore, these force-feedback devices may be embedded (i.e., available in appropriate form factor and reasonably priced) in ROV consoles for both training and performance enhancement.

At least two classes of information can be displayed to the ROV pilot trainee using force-feedback: (1) information about the ROV's immediate dynamic state (e.g., velocity, acceleration, thrust, hydrodynamic vehicle drag, tether drag, currents, etc.), and (2) information concerning the ROV's deviation from some desired or ideal dynamic state (i.e., errors between the immediate dynamic state and some idealized dynamic state). Initially, we are focusing on how the feedback of the ROV's immediate dynamic state using force affects training transfer.

We are training ROV dynamic position/orientation control in simple tracking scenarios. In these scenarios, two general situations are encountered: (1) the ROV acquires and maintains position/orientation with respect to a stationary target (Tracking Task 1), and (2) the ROV maintains position/orientation with respect to a moving target (Tracking Task 2). These scenarios exercise and demonstrate very basic ROV control skills. The first tracking task is representative of situations in which a pilot is attempting to keep an object or feature within the ROV's sensor envelope and, potentially, in which the ROV is being subjected to external forces (e.g., currents). Real world examples of the second tracking task include ROV applications such as: (1) swimmer detection and response, and (2) pipe following and inspection under low-visibility conditions. These tracking tasks also lend themselves to meaningful experimental metrics and analysis.

During the first year of our efforts, we explored the effects of force-feedback on dynamic position/orientation control in environments where the degrees-of-freedom (DOFs) of the ROV have been constrained to 1-DOF. Constraining the degrees-of-freedom allows us to tightly control our experiments (i.e., thereby aiding analysis and understanding of the situation and results) and to keep our experimental apparatus simple. Furthermore, positive training transfer in these simplified environments is quite likely to extend to more complex environments. The three experimental cases which we

explored during the first year of our effort were: (1) tracking a target which moved discretely (Tracking Task 1) with no external environmental influences, (2) tracking a target which moved continuously (Tracking Task 2) with no external influences, and (3) tracking a nonmoving target (Tracking Task 1) under the influence of a varying current.

ROV Path Awareness

In order to aid in general navigation and general situation awareness, the pilot of an ROV must be aware of current position and orientation in some world reference frame. In addition, the ROV pilot must be cognizant of relevant portions of the ROV's position/orientation history. (We will refer to the history and current position/orientation of the ROV as the "path"). Path awareness is especially critical for tether management and to avoid tether entanglement. The pilot's awareness of the ROV's path is based upon control inputs to the ROV and sensor feedback from the ROV. This feedback is highly impoverished both in fidelity and form and, typically, does not include any significant sensorimotor feedback. An ROV pilot's skill in understanding and recalling the ROV's path through space, therefore, is an important and non-trivial skill required for effective ROV piloting.

Training the path awareness skill seems to require: (1) making the trainee aware that such a skill is important, (2) practicing so that the skill may be developed, and (3) performing appropriate training interventions so that 1 and/or 2 may be more efficiently achieved. Component skills include the ability to judge and remember the length of any given segment in the path as well as the overall relationships between path segments. Pragmatic measures of expertise include: (1) the ability to estimate the ROV's current position and orientation in three-space (in either absolute or relative coordinates), and (2) the ability to recall the entire path through three space by either taking the same path in the future or by returning along the same path to the missions starting location. It should be noted that path awareness is a required skill for ROV pilots and is also a major component in the effective navigation of synthetic and remote spaces in general. In our experiments, we are exploring the efficacy of certain forms of sensorimotor cues and interventions in helping develop path awareness skills.

Two general classes of sensorimotor cues and interventions will be considered: (1) cues based on absolute position feedback, and (2) cues based on "effort". Absolute position cues are generated using haptic maps of the remote or virtual environment. Cues based on effort (i.e., work, energy expendi-

ture, or some similar measure), establish some scaled relationship between effort expended by the pilot on controlling the ROV (i.e., where controls have been constructed to require effort) and the distance traveled by the ROV in space.

During the first year, we initiated research to explore the role that sensorimotor cues and interventions may have on training transfer for distance judgment in simple 1-DOF motion scenarios. Conceptually, the task consists of subjects using a simulated ROV to search for a given number of "objects" along the 1-DOF path. Objects will be seen only when the ROV is very close to them. Subjects' ability to accurately judge the inter-object path segments will be tested in both "replay" and "return" scenarios. In the control case, trainees will use a 1-DOF, spring-return-to-center, joystick to control a simulated ROV which is used to "learn" the simulated remote environment. Experimental subjects are trained under the same conditions as the control group except that additional sensorimotor cues are provided. Conditions to be tested include both absolute position-based and effort-based sensorimotor cues.

1.12.4 Program Management and Web Support

Software dependencies throughout the TRANSOM project have been identified and modified so that the TRANSOM system is portable to other machines. In addition, all project source software has been consolidated into manageably sized modules. A source code cross-reference program was installed, and a project Web site was developed, including a source code browser. A WEB-based e-mail archiving system has been installed, so that any TRANSOM team member can now address e-mail to this archive so that a log of R&D discussions will be maintained as a searchable, browsable database. TRANSOM software has been brought under source code control using the CVS system. A regular schedule for building weekly internal MIT/BBN distribution packages and less frequent team distributions for Imetrix and McDonnell Douglas was established, and internal and team distributions were generated. These TRANSOM software distribution packages provide for turnkey installation on remote sites, so that the most recent version of the TRANSOM VE

system can be easily distributed, installed and made available to all TRANSOM researchers. The TRANSOM public web page has also been installed (URL: <http://mimsy.mit.edu/Transom/Public-page/project.html>). This page is under construction and will be updated frequently.

1.12.5 Publication

Pioch, N., B. Roberts, and D. Zeltzer. "A Virtual Environment for Learning to Pilot Remotely Operated Vehicles." Submitted to *Virtual Syst. Multimedia*.

1.13 Virtual Environment Technology for Training: Development of Haptic Interfaces

Sponsor

U.S. Navy - Office of Naval Research/
Naval Air Warfare Center
Contract N61339-96-K-0002

Project Staff

Dr. Mandayam A. Srinivasan, Dr. J. Kenneth Salisbury, Dr. Cagatay Basdogan, Dr. G. Lee Beauregard, David E. DiFranco, Chih-Hao Ho, I-Chun A. Hou

1.13.1 Overview

Haptic displays are emerging as effective interaction aids for improving the realism of virtual worlds. Being able to touch, feel, and manipulate objects in virtual environments have a large number of exciting applications. The underlying technology, both in terms of electromechanical hardware and computer software, is becoming mature and has opened up novel and interesting research areas.

Over the past few years, we have developed device hardware, interaction software and psychophysical experiments pertaining to haptic interactions with virtual environments (general reviews on haptic interfaces can be found in Srinivasan, 1994; 1995).⁴⁵ Two major devices for performing psychophysical experiments, the linear and planar

⁴⁵ M.A. Srinivasan, "Virtual Haptic Environments: Facts Behind the Fiction," *Proceedings of the Eighth Yale Workshop on Adaptive and Learning Systems*, Center for Systems Science, Yale University, New Haven, June 1994; Srinivasan M.A., "Haptic Interfaces," in *Virtual Reality: Scientific and Technical Challenges*, eds. N.I. Durlach and A.S. Mavor, Report of the Committee on Virtual Reality Research and Development, (Washington, D.C.: National Research Council, National Academy Press, 1995); M.A. Srinivasan,

graspers, have been developed. The linear grasper is capable of simulating fundamental mechanical properties of objects such as compliance, viscosity and mass during haptic interactions. Virtual walls and corners were simulated using the planar grasper, in addition to the simulation of two springs within its workspace. The PHANTom, another haptic display device developed previously by Dr. Salisbury's group at the MIT Artificial Intelligence Laboratory, has been used to prototype a wide range of force-based haptic display primitives. A variety of haptic rendering algorithms for displaying the shape, compliance, texture, and friction of solid surfaces have been implemented on the PHANTom. All the three devices have been used to perform psychophysical experiments aimed at characterizing the sensorimotor abilities of the human user and the effectiveness of computationally efficient rendering algorithms in conveying the desired object properties to the human user.

The following sections summarize the progress over the past year in our "Touch Lab" at RLE. We mainly describe the major advances in a new discipline, *Computer Haptics* (analogous to computer graphics), that is concerned with the techniques and processes associated with generating and displaying haptic stimuli to the human user.⁴⁶

1.13.2 Haptic Display of 3D Objects in Virtual Environments: A Ray-based Haptic Interaction Technique

We have developed a new ray-based haptic rendering technique for haptic display (rendering) of 3D rigid objects in virtual environments. A set of software algorithms have been developed and named as the "Haptic-Binder".⁴⁷ The binder holds the virtual objects together and works with a force reflecting haptic interface to enable the user to touch and feel arbitrary 3D polyhedral virtual objects. The rendering model has four major components: a hierarchical database for storing geometrical and material properties of objects, collision detection algorithms, a simple mechanistic model for computing the forces of interaction

between the 3D virtual objects and the force-reflecting device, and a haptic filtering technique for simulating surface details of objects such as smoothness and texture.

In the past, Zilles and Salisbury⁴⁸ developed a constraint-based method to render generic polygonal meshes. They defined a new point, namely the "god-object," to represent the location of the ideal haptic interface point. They used Lagrange multipliers to compute the new location of the god-object point. We propose a new haptic rendering method that utilizes the ray tracing techniques of computer graphics for detecting collisions between 3D polygonal objects (rigid) and the generic stylus of the haptic device. The proposed method is more suitable for simulating "tool-object" interactions than earlier techniques. The stylus is modeled as a line segment whose position and orientation are provided by the encoder signals in the haptic interface. We display the graphical model of the simulated stylus and update its tip and tail coordinates as the user manipulates the actual one (see figure 12a), detect any collisions between the simulated stylus and the virtual object, estimate the reaction force, and finally reflect this force to the user via the haptic device.

We first load the graphical models of 3D objects, made of triangular elements, into the computer from a data file to construct a database for material and geometrical properties of objects. The information in this database includes the indices of triangular elements, coordinates of vertices of each triangular element and the connectivity of triangular elements in a global coordinate system. Using the information in the database, algorithms compute the bounding box of each object (i.e., a box that is constructed from maximum and minimum global coordinates of the object along the x, y, and z axes) and the bounding box and surface normal of each triangular element of the object. As the user manipulates the stylus of the force feedback device, we check for collisions between the simulated stylus and the virtual objects in the environment (figure 12a). The detection of collisions between simulated stylus and 3D virtual objects occur in three consecutive stages:

"Haptic Interfaces: Hardware, Software, and Human Performance," *Proceedings of the NASA Workshop on Human-computer Interaction and Virtual Environments*, NASA Conference Publication 3320, 1995, pp. 103-121.

⁴⁶ M.A. Srinivasan and C. Basdogan, "Haptics in Virtual Environments: Taxonomy, Research Status, and Challenges," *Computers and Graphics* 21(4): (1997).

⁴⁷ C. Basdogan, C. Ho, and M.A. Srinivasan, "Haptic Display of 3D Rigid Objects in Virtual Environments," accepted for presentation at the ASME Winter Annual Meeting, Dallas, Texas, 1997.

⁴⁸ C.B. Zilles and J.K. Salisbury, "A Constraint-based God-object Method for Haptic Display," IEEE International Conference on Intelligent Robots and System, 1995.

1. Collision detection between the simulated stylus and the bounding box of the virtual object
2. Collision detection between the simulated stylus and the bounding box of each triangular element
3. Collision detection between the simulated stylus and the triangular element itself

Following the detection of collision, we compute the distance between the collision point (i.e., ideal haptic interface point) and the tip of the stylus (i.e., virtual haptic interface point) and multiply this distance with the gross impedance of the object to compute the total reaction force. Then, for frictionless objects, the component of this force along the direction of surface normal is calculated and sent as a force command to the force feedback device to provide the user with the tactual sensation of object shapes (see figure 12b).

1.13.3 Haptic Display of Surface Details of 3D Objects in Virtual Environments

We have developed techniques that enable the user to manually feel the surface properties (e.g., smoothness and texture) of objects in virtual environments. Our research on haptic display of surface details can be summarized into two main groups: (1) haptic smoothing of object surfaces, and (2) rendering of haptic textures. To convey to the user the tactual feeling of smooth surfaces, we compute the force vector for each vertex and smoothly interpolate its direction over the polygonal surfaces. In order to display rough surfaces haptically, we either perturb the force vector using the local gradient of texture field (force perturbation) or divide the surface of the object into smaller polygons to change its topology (displacement mapping).

Haptic Smoothing of Object Surfaces

When smooth and continuous object shapes are approximated by polyhedra for haptic rendering, the user does not perceive the intended shape. Instead, the discrete edges between polygons as well as the planar faces of the polygons are felt. To minimize such undesirable effects,

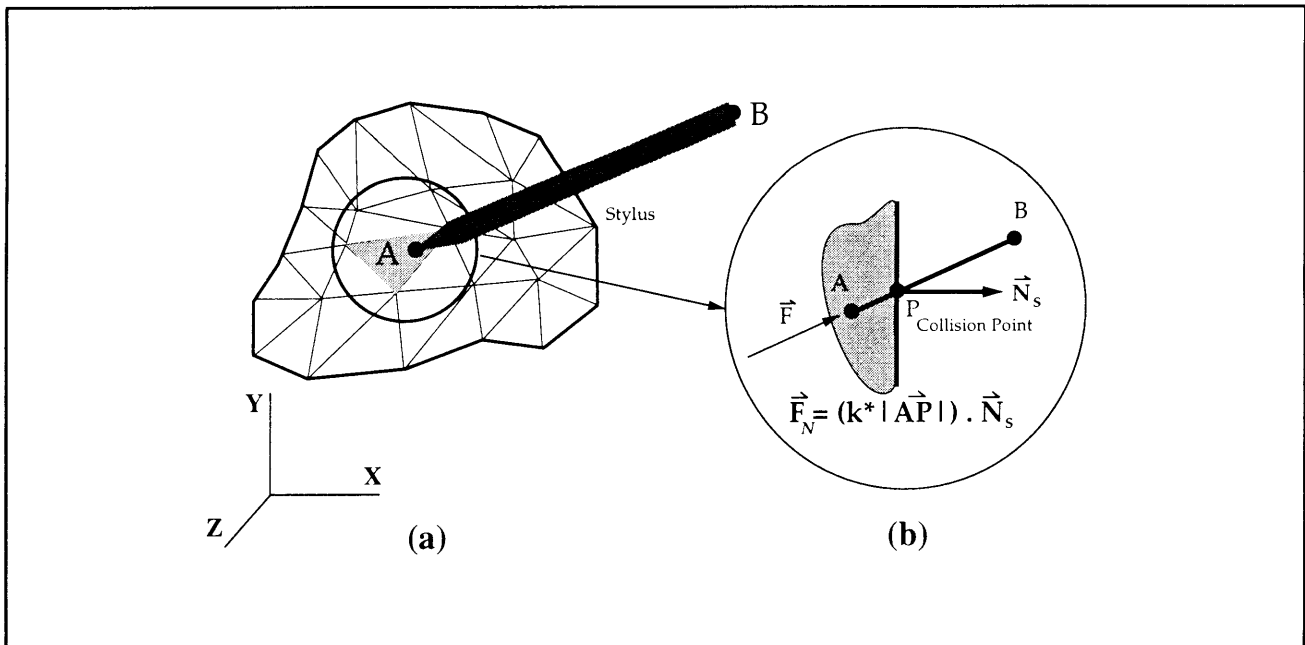


Figure 12. (a) The stylus is modeled as a line segment (\overline{AB}) for the purposes of fast collision computations. (b) Following the detection of collision, the reaction (contact) force is computed using a mechanistic model ($\vec{F} = \kappa \vec{AP}$).

Morgenbesser and Srinivasan⁴⁹ proposed force shading (see figure 13a). In this method, which falls within the general class of force mapping techniques, the force vector is interpolated over the polygonal surfaces such that its direction varies continuously. Consequently, the surfaces of virtual objects feel smoother than their original polyhedral representations. This technique is analogous to Phong shading in computer graphics.

Rendering of Haptic Textures

Visual rendering of surface details such as textures and bumps, to add realism to the appearance of the 3D objects, has been a challenging research topic because of its variety and complexity. In multimodal VEs, the computational burden is further increased because of the need to represent and synchronously display the macro and micro surface details through touch and/or sound. Minsky et al.⁵⁰ implemented and performed perceptual experiments on a variety of textures with a 2D haptic display. We developed two haptic rendering techniques that are similar to those used in computer graphics, and are suitable for 3D haptic displays: (1) force perturbation, and (2) displacement mapping.

Force perturbation is a technique of modifying the direction and magnitude of the force vector to generate surface effects such as roughness (figure 13b). Analogous to the "bump mapping" technique of computer graphics,⁵¹ force perturbation provides the user with a sense of surface details such as textures and bumps. To generate visual bumps on the surface of the object, Max and Becker⁵² developed a formulation which is based on the original surface normal and the local gradient of the height field, as follows:

$$\vec{M} = \vec{N} - \vec{\nabla}h + (\vec{\nabla}h \cdot \vec{N})\vec{N}$$

where, \vec{M} is the perturbed surface normal, \vec{N} is the original surface normal, and $\vec{\nabla}h$ is the gradient of the texture height field. We used the same approach and perturbed the direction and magnitude of the force vector to generate bumpy or textured surfaces that can be sensed tactually by the user.⁵³

In computer graphics displacement mapping technique, the actual geometry of the object is modified to visually display the surface details.⁵⁴ In order to generate microtextures, the geometry of the surface

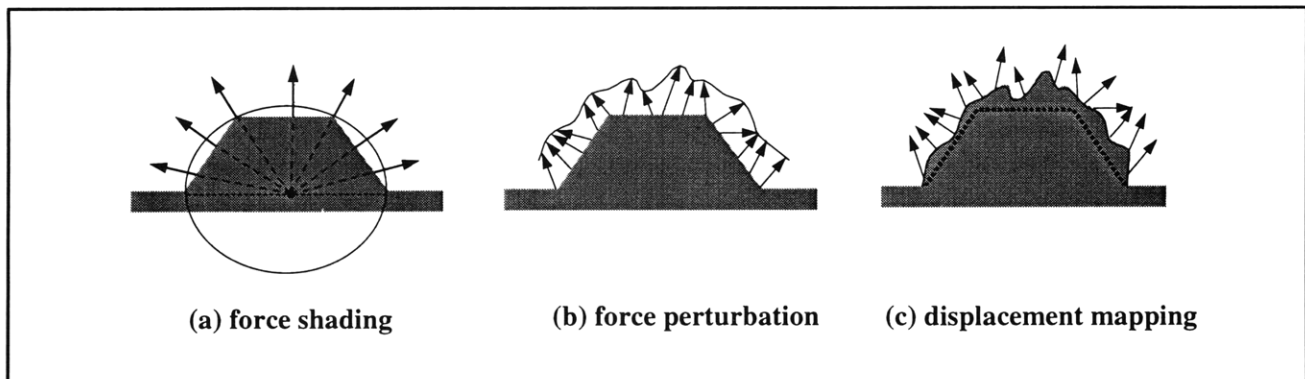


Figure 13. Haptic rendering techniques for simulating surface details in virtual environments. The arrows represent the reflected force vectors. In each case, the shaded area represents the displayed geometry of the object and the dark boundary line represents the geometry of the surface intended to be haptically perceived by the user.

⁴⁹ H.B. Morgenbesser and M.A. Srinivasan, "Force Shading for Haptic Shape Perception," *Proceedings of the ASME Dynamic Systems and Control Division*, DSC-Vol. 58, 1996, pp. 407-412.

⁵⁰ M.M. Minsky, M. Ouh-Young, O. Steele, F. Brooks, and M. Behensky, "Feeling and Seeing: Issues in Force Display," *Proceedings of Symposium on Interactive 3D Graphics*, ACM Siggraph, Snowbird, Utah, 1990, pp. 235-243.

⁵¹ J.F. Blinn, "Simulation of Wrinkled Surfaces," *ACM Proceedings of SIGGRAPH* 12(3): 286-292 (1978).

⁵² N.L. Max and B.G. Becker, "Bump Shading for Volume Textures," *IEEE Computer Graphics and App.* 14: 18-20, 1994.

⁵³ C. Basdogan, C. Ho, and M.A. Srinivasan, "Haptic Display of 3D Rigid Objects in Virtual Environments," submitted to the ASME Winter Annual Meeting, 1997.

⁵⁴ J.D. Foley, A. van Dam, S.K. Feiner, and J.F. Hughes, "Computer Graphics: Principles and Practice," Addison-Wesley, 1995.

has to be composed of tiny polygons so that its surface can be modified point by point. During the rendering of new surface, normals need to be updated in the geometrical database. We observed that detecting collisions or locating the new position of the surface point each time the generic probe is moved creates ambiguities due to the existence of fine details on the surface geometry. Force discontinuities occur as the user strokes the surface of object with the probe to feel the textures displaying haptic stimuli to the human user.⁵⁵

Modeling of haptic textures require an understanding of textures that exist in nature and their interactions with the human tactile sensory system. We know that textures in nature come in many varieties, hence no single model can represent all the textures. Various graphical techniques have been developed in the past to generate photorealistic textured objects.⁵⁶ In our studies, we first focused on frequency and height as the two major indicators of the haptic textures. We then classified the haptic texturing techniques into two groups, as in computer graphics: (1) image-based haptic texturing, and (2) procedural haptic texturing.

Image-based Haptic Texturing

In the displacement mapping used in computer graphics, the images of textures are wrapped around 3D objects to make them look more realistic. In haptic texturing, the aim is to make the user feel the textures of the visual image that has already been mapped onto the 3D surface. To simulate image-based textures, we construct a texture field from the 2D image data. Assuming that gray scale intensities of 2D image can be used directly as height indicators to generate a haptic texture field, we can associate each texture coordinate (u,v) with the coordinate of each vertex (x,y,z). We then map these heights onto the object surface using the two-stage texture mapping techniques used in computer graphics.⁵⁷ For haptic rendering, we compute the local gradient of the height field at the collision point and perturb the force vector

direction accordingly.⁵⁸ Obviously, this method of associating heights to images is neither unique nor fool-proof.

Procedural Haptic Texturing

The aim of procedural haptic texturing is to generate synthetic texture fields using mathematical functions for the height field. Once the function that describes the texture field is known, gradient vector (∇h) at the point of contact can be easily calculated either to perturb the reaction force vector or to modify the surface geometry of the object as in displacement mapping. We have generated a variety of synthetic haptic textures by using various procedural texturing techniques of computer graphics in conjunction with force perturbation.⁵⁸ Examples of formulations we have used are based on Fourier series, stochastic functions such as the filtered noise, and fractals.

1.13.4 Visual-Haptic Interactions

The development of a visual and haptic rendering program and the investigations into human visual-haptic interactions are described in this section. We developed a tool-kit (The MAGIC Toolkit) that is composed of an application program and a library file that allow users to see, manually feel, create, edit, and manipulate objects in a virtual environment.⁵⁹ Using a force reflecting haptic interface, such as the PHANTOM, a user can build a complex virtual object or scene by adding object primitives to the virtual workspace. Object primitives, such as cylinders and spheres, are pre-programmed and have both visual and haptic characteristics that can be modified with a touch to the virtual menu wall.

Using the MAGIC Tool-kit, we have conducted human experiments to investigate the role of visual-haptic size ratio, the visual and haptic sensory feedback in isolation or together, the effects of visual scaling on training, and various cursor control paradigms.⁵⁹ The task was to navigate through a

⁵⁵ M.A. Srinivasan and C. Basdogan, "Haptics in Virtual Environments: Taxonomy, Research Status, and Challenges," *Computers and Graphics*, 21(4): (1997).

⁵⁶ D.S. Ebert, F.K. Musgrave, D. Peachey, K. Perlin, and S. Worley, "Texturing and Modeling," AP Professional, Cambridge, Massachusetts, 1994.

⁵⁷ E.A. Bier and K.R. Sloan, "Two-Part Texture Mapping," *IEEE Computer Graphics and Applications*, pp. 40-53, 1986.

⁵⁸ C. Basdogan, C. Ho, and M.A. Srinivasan, "Haptic Display of 3D Rigid Objects in Virtual Environments," accepted for presentation at the ASME Winter Annual Meeting, Dallas, Texas, 1997.

⁵⁹ I.A. Hou and M.A. Srinivasan, "Multimodal Virtual Environments: The MAGIC Toolkit and Visual-Haptic Interaction Paradigms," submitted to the ASME Winter Annual Meeting, 1997.

maze with visual and/or haptic feedback. Results show that subjects preferred large visual-haptic ratios, small haptic workspaces, and a position controlled cursor. Subjects performed best with a large visual display and a small haptic workspace. Also, subjects performed best when both visual and haptic cues were given, with a slight decrease in performance when only haptic cues were given, and with a significant decrease in performance when only visual cues were given. The performance of the subjects improved linearly with increases in visual display size, when subjects were initially trained on the largest visual size in the series. Among various cursor control methods, subjects performed best when there was a high correlation in cursor position and movement between the visual and haptic workspaces.

1.13.5 Haptics in Distributed Virtual Environments

In order to make haptics and our research studies accessible and transferable to the others, we opted to integrate haptics into the Web. A demonstration version of the visual-haptic experiment as described in section 1.10 (ERHO section) using the PHANToM haptic interface was developed to be used across the World-Wide-Web. The program was written in Java, using multithreading to create separate visual and haptic control loops, thereby increasing the speed of the haptics loop to keep the program stable despite its graphics overhead. The application program has been placed on the Laboratory for Human and Machine Haptics web page (<http://touchlab.mit.edu>), to be executed by any remote user with a PHANToM and a Windows NT computer running Netscape for WWW access. Remote users can download a dynamic link library and some Java classes from the web page to their computer, and then run the program in their web browser. Users are asked to discriminate the stiffness of sets of two springs, displayed visually on the screen and haptically with the PHANToM, and to send in their responses via an e-mail window in the web page. Thus, we now have the ability to perform perceptual experiments with multi-modal VEs across the internet.

1.13.6 Publications

Basdogan, C., C. Ho, and M.A. Srinivasan. "Haptic Display of 3D Rigid Objects in Virtual Environments." Submitted to the ASME Winter Annual Meeting, 1997.

Beauregard, G.L., and M.A. Srinivasan, *Sensori-*

motor Interactions in the Haptic Perception of Virtual Objects. RLE TR-607. MIT, 1996.

Hou, I.A., and M.A. Srinivasan. "Multimodal Virtual Environments: The MAGIC Toolkit and Visual-Haptic Interaction Paradigms." Submitted to the ASME Winter Annual Meeting, Dallas, Texas, 1997.

Morgenbesser, H.B., and M.A. Srinivasan. *Force-Shading for Shape Perception in Haptic Virtual Environments*. RLE TR-606. MIT, 1996.

Srinivasan, M.A., and C. Basdogan. "Haptics in Virtual Environments: Taxonomy, Research Status, and Challenges." Submitted to *Computers and Graphics* 21(4): (1997).

1.14 Role of Skin Biomechanics in Mechanoreceptor Response

Sponsor

National Institutes of Health
Grant R01-NS33778

Project Staff

Dr. Mandayam A. Srinivasan, Dylan J. Birtolo, Suvranu De, Amanda S. Myers, Balasundar I. Raju, Kimberly J. Voss

1.14.1 Overview

When we touch an object, the source of all tactile information is the spatio-temporal distribution of mechanical loads on the skin at the contact interface. The relationship between these loads and the resulting stresses and strains at the mechanoreceptive nerve terminals within the skin plays a fundamental role in the neural coding of tactile information. Although empirical determination of the stress or strain state of a mechanoreceptor is not possible at present, mechanistic models of the skin and subcutaneous tissues enable generation of testable hypotheses on skin deformations and associated peripheral neural responses. Verification of the hypotheses can then be accomplished by comparing the calculated results from the models with biomechanical data on the deformation of skin and subcutaneous tissues, and neurophysiological data from recordings of the responses of single neural fibers. The research under this grant is directed towards applying analytical and computational mechanics to analyze the biomechanical aspects of touch—the mechanics of contact, the transmission of the mechanical signals through the skin, and

their transduction into neural impulses by the mechanoreceptors.

1.14.2 Determination of the Shape and Compressibility of the Primate Fingertip (Distal Phalanx)

The first step in performing mechanistic analyses of the primate fingertip is to determine its geometric and material properties. The three-dimensional (3D) external geometry of the primate fingertips was determined from accurate epoxy replicas of human and monkey fingertips. Using a videomicroscopy setup, we obtained images of orthographic projections of the epoxy replicas at various known orientations. The images were then digitized and processed to determine the boundary of the finger at each orientation. By combining the boundary data for all the different orientations, we were able to reconstruct the 3D external geometry of the fingertip.⁶⁰ We have reconstructed several human and monkey fingertips by using this method.

For mechanistic modeling of the human fingerpad, the Poisson's ratio, which is a measure of its compressibility, is required as an input to the mathematical models. The Poisson's ratio for the human fingerpad *in vivo* is unknown at present. In previous noninvasive experiments on human subjects, we have measured the change in volume of the fingerpad under static indentations with different indentors.⁶¹ Our results show that the compressibility of the fingertip increases with increases in both the depth of indentation and the contact area with the indenter. The highest change in fingertip volume was about 5 percent. We have also developed an experimental setup involving a computer controlled linear actuator for fingertip volume change measurements under dynamic conditions.⁶² The results show that reductions in fingertip volume are in phase with stimulus variations, with an increase in their mean value over time. The volume

changes during the ramp phase increase linearly with indenter displacement and are independent of velocity; during saw tooth stimulations, however, the nature of the hysteresis loops depend on velocity of indentation.

1.14.3 Fingertip Models and Finite Element Analysis

We have performed linear and nonlinear finite element analysis of a series of mechanistic models of the fingerpad under a variety of mechanical stimuli.⁶³ The models range from a semi-infinite medium to a 3D model based on the actual finger geometry, and composed of a homogeneous elastic material, a thick elastic shell containing a fluid or a multilayered medium. Simulations of the mechanistic aspects of neurophysiological experiments involving mapping of receptive fields with single point loads, determination of spatial resolution of two-point stimuli, and indentations by single bars as well as periodic and aperiodic gratings have been carried out for the 2D and 3D models. We have also solved the nonlinear contact problem of indentations by cylindrical objects, rectangular bars and sinusoidal step shapes. The large number of numerical calculations needed even for the linear two dimensional models necessitated the use of the Cray-C90 at the NSF Pittsburgh Supercomputer Center.

The results show that the model geometry has a significant influence on the spatial distribution of the mechanical signals, and that the elastic medium acts like a low-pass filter in causing blurring of the mechanical signals imposed at the surface. Multilayered 3D models of monkey and human fingertips accurately predicted the surface deformations under a line load, experimentally observed by Srinivasan.⁶⁴ The same models predicted the experimentally observed surface deformations under cylindrical indentors, as well. These 3D finite

⁶⁰ T.R.R. Perez, K. Dandekar, and M.A. Srinivasan, "Videomicroscopic Reconstruction of the Human Finger," Project Report to the MIT Summer Science Research Program, 1992.

⁶¹ M.A. Srinivasan, R.J. Gulati, and K. Dandekar, "In Vivo Compressibility of the Human Fingertip," *Advances in Bioengineering*, ASME, Vol. 22, 1992, pp. 573-576.

⁶² W.E. Babiec, *In Vivo Volume Changes of the Human Fingerpad Under Indentors*, B.S. thesis, Dept. Mech. Eng., MIT, 1994.

⁶³ M.A. Srinivasan and K. Dandekar, "Role of Fingertip Geometry in the Transmission of Tactile Mechanical Signals," *Advances in Bioengineering*, ASME, 22: 569-572 1992; M.A. Srinivasan and K. Dandekar, "An Investigation of the Mechanics of Tactile Sense Using Two Dimensional Models of the Primate Fingertip," *J. Biomech. Eng.* 118: 48-55 1996; K. Dandekar and M.A. Srinivasan, "Tactile Coding of Object Curvature by Slowly Adapting Mechanoreceptors," In *Advances in Bioengineering*, ed. M.J. Askew, Chicago: ASME, BED-28: 41-42 1994; K. Dandekar and M.A. Srinivasan, *Role of Mechanics in Tactile Sensing of Shape*, RLE TR-604, MIT, 1997.

⁶⁴ M.A. Srinivasan, "Surface Deflection of Primate Fingertip Under Line Load," *J. Biomech.* 22(4): 343-349 (1989).

element models were used to simulate neurophysiological experiments involving indentation by rectangular bars, aperiodic gratings, cylindrical indentors, and sinusoidal step shapes. Since our earlier model had less resolution than what was required for simulation of indentation by objects of steep curvature, we developed a 3D finite element model of the fingertip with a highly improved resolution of 150 microns. The entire model had about 90,000 degrees of freedom. About 20 mechanical quantities (normal stresses, shear stresses, normal strains, stress and strain invariants, principal stresses and principal strains) were calculated through simulations in order to be matched with experimentally recorded neural responses. These simulations have provided a hypothesis for the link between the mechanical stress/strain measures and the experimentally recorded neural impulses emanating from the slowly adapting receptors. Strain energy density distribution in the vicinity of the mechanoreceptors was able to predict the shape of the spatial response profile for step shapes of varying steepness⁶⁵ as well as rectangular bars. This confirms our earlier hypothesis and also extends its validity to the case of indentation by steeply curved objects such as the edges of rectangular bars.

1.14.4 Incorporation of Fingerpad Dynamics

In order to model the dynamic behavior of the fingerpad, viscoelasticity has been incorporated into our FEM model. To this end, the biomechanical data obtained from the indentation of the fingerpads of several human subjects using different indenter geometries was utilized. A consistent normalization scheme was developed which showed that most of the variation in the data obtained from various subjects was scalable by a single parameter. We then came up with a second order Kelvin model which satisfactorily explains much of the observed data for a truncated conical indenter. The Correspondence Principle was invoked to extend these results to obtain the material parameters of a generalized 3D linear viscoelastic solid. These parameters were then incorporated into a 2D plane strain and a 3D layered FEM model. The results obtained from the FEM models match the observed force-displacement data for all the indentors (truncated conical, cylindrical and flat plate indentors) used in the original biomechanical experiments.

We have also come up with a novel modeling paradigm for predicting the mechanoreceptor responses which are transmitted to the central nervous system via the peripheral nerves. We have proposed to decouple the tissue dynamics from neural dynamics using known experimental results. We have characterized the tissue dynamics quite satisfactorily. The "overall dynamics" relating the input sinusoidal loading and the observed neural responses from each class of mechanoreceptors has also been modeled. From these we have obtained the dynamics of each class of mechanoreceptors. We are now in a position to predict the spatial response profiles observed during the stroking of complex shapes (toroids, wavy surfaces, and sinusoidal step shapes). These simulations are under way and have produced satisfactory results. The next step is to use the same models (tissue mechanics and receptor transduction models) to predict population response profiles, the "tactile image" that the peripheral nervous system obtains from a shape indenting the compliant fingerpad skin.

1.14.5 MRI Data for Realistic Internal Geometry and Deformation

A further improvement in our models requires the incorporation of internal geometry of the fingerpad. We have used Magnetic Resonance Imaging (MRI) techniques to obtain cross-sectional images of the fingerpad with a resolution of 60-100 μm X 60-100 μm . Geometric surfaces which simulate edges and surfaces of objects such as a line load, various sized rectangular bars and cylinders were used to load the fingertip. We designed a custom probe which allows for the application of various static loads onto the fingerpad. The load was applied to the finger at a 30 degree angle from the nailbed to simulate the normal orientation of the fingerpad during touch. Digital image processing techniques were used to filter the images and to detect the edges of the tissues located in the fingerpad. Edge detection algorithms such as conformable contours ("snakes") were implemented to determine the location of tissue layers and their displacement. Published data on histology, cytology and anatomy were used to identify each structure in the fingertip. From this data, the biomechanical behavior of the tissue is being determined and will be used to verify performance of the computational models.

⁶⁵ M.A. Srinivasan, and R.H. LaMotte, "Tactile Discrimination of Shape: Responses of Slowly and Rapidly Adapting Mechanoreceptive Afferents to a Step Indented into the Monkey Fingerpad," *J. Neuroscience* 7: 1682-1697 (1987).

1.14.6 Tactile Stimulator for Biomechanical Experiments

Currently, we are building a one-dimensional tactile stimulator that will allow us to further characterize the biomechanical properties of the fingerpad. This device has been designed to deliver dynamic indentation stimuli to the human fingerpad either in the force control mode or in displacement control mode. Various indenter shapes can be attached to the stimulator to provide sinusoidal inputs with frequencies ranging from 10 to 300 Hz. Results from these experiments will be used to extend the biomechanical behavior of our models of the fingerpad.

1.14.7 Publications

Dandekar, K., and M.A. Srinivasan. *Role of Mechanics in Tactile Sensing of Shape*. RLE TR-604, MIT, 1997.

Srinivasan, M.A., and K. Dandekar. "An Investigation of the Mechanics of Tactile Sense Using Two-Dimensional Models of the Primate Fingertip." *J. Biomech. Eng.* 118: 48-55 (1996).

1.15 Human and Robot Hands: Mechanics, Sensorimotor Functions and Cognition

Sponsor

U.S. Navy - Office of Naval Research
Grant N00014-92-J-1844

Project Staff

Dr. Mandayam A. Srinivasan, Dr. J. Kenneth Salisbury, Dr. Robert H. LaMotte,⁶⁶ Dr. Robert D. Howe,⁶⁷ Jyh-Shing Chen, Steingrimur P. Karason, Jung-Chi Liao

1.15.1 Overview

The premise of this project is that the integrated study of human and robot haptics can provide complementary knowledge of the processes of prehension and manipulation. From the human side we wish to understand the basic mechanical, perceptual and strategic capabilities that lead to the dexterity and deftness we observe in human task performance. By studying the underlying competences that humans bring to bear on task performance, we seek guidelines on how to better build robots. From the robotic side we wish to understand how mechanism and sensor design choices can best be made to maximize grasping and manipulative competences. By better understanding the mechanical demands of task performance, we seek to understand the performance demands which underlie skillful human manipulation.

The main components of the research conducted under this project are (1) development of new hardware for robotic and human studies, (2) processing of robot sensor signals and task-level control of the devices, (3) investigation of the mechanics of human fingerpad-object contact, (4) experiments on human perception and control of forces using some of the devices, and (5) development of a computational theory of haptics that is applicable to both robots and humans. The subsections to follow provide descriptions of the results obtained mainly by the Human Haptics Group at the Touch Lab in RLE.

1.15.2 Human Haptic Interaction with Soft Objects

The haptic interaction of humans with soft objects was studied from three perspectives: (1) softness discrimination, (2) force control, and (3) contact visualization.⁶⁸ The abilities of humans in actively discriminating softness was measured by using a specimen presenter system which was built to randomly present the specimens. Two experimental paradigms (one- and two-finger) and three-finger conditions (normal, finger cot, and rigid thimble) were used to examine the importance of various sources of information. When using one-finger, the just noticeable difference (JND) was about 5

⁶⁶ Yale University School of Medicine, New Haven, Connecticut.

⁶⁷ Harvard University, Cambridge, Massachusetts.

⁶⁸ J.-S. Chen, *Human Haptic Interaction with Soft Objects: Discriminability, Force Control, and Contact Visualization*. Ph.D. diss., Dept. of Mech. Eng., MIT, 1996.

percent for normal and finger cot conditions and increased to about 50 percent with the thimble. The JND results from two-finger discrimination were lower than the one-finger results for both normal and finger cot conditions, but were higher with the thimble. Examination of the forces exerted on the specimens during two-finger discriminations revealed possible underlying discrimination strategies.

In the force control experiments, subjects were asked to exert several levels of constant forces under various finger conditions. The results indicated that the errors from tracking with visual feedback was significantly lower than without visual feedback. No significant differences in force control were found with either different softnesses or under the three-finger conditions. The absolute errors were higher when controlling higher target forces. Significant difference in force control was found between the two hands of the subject who showed handedness in the softness discrimination experiments.

For contact visualization, a real-time imaging setup was built which consisted of a videomicroscopy system and a tactile stimulator system. By using this setup, real-time images from the contact region as well as the contact forces were digitized. Various image processing techniques were developed and applied in order to analyze and improve the contact images to distinguish between the contact and noncontact regions. Contact variables such as force, nominal contact area, actual contact area, nominal pressure, and actual pressure were analyzed. Based on the visualization of the active slip phenomenon, a hypothesis on the contact pressure distribution was proposed. The developed hardware can be used for evaluating human haptic abilities and providing biomechanical visualizations. The results obtained have implications on the mechanisms of contact and find applications in dextrous robot finger and haptic interface designs as well as in various steps involved in the automatic fingerprint identification systems.

1.15.3 Frictional Properties of the Human Fingertpad

A 2-DOF robot has been developed to serve as a high-precision tactile stimulator with a two-axis force sensor mounted on it. Flat plate indentors are attached to the force sensor to indent and stroke the fingertpad of the subject's index finger. The tactile stimulator is capable of applying various dynamic stimuli such as different indentation depths and stroke velocities. During the stroking process, in addition to the displacements along the horizontal

and vertical directions, the normal and shear forces on the fingertpad are measured by force sensors at a sampling rate of 500 Hz. A 80486 computer controls the motion of the tactile stimulator and records the force data.

A videomicroscopy system has been set up to capture the image sequences of the fingertpad while stroking. This system consists of a high-resolution CCD camera, video zoom lens, a fiber optic light source, and a frame grabber hosted by another 80486 computer. During the experiments, this computer is operated synchronously with another 80486 computer that controls the tactile stimulator and records data from force sensors. This synchronization is accomplished by a timer board and digital I/O setup, which allows us to match the object-fingertpad contact images with corresponding force data.

Three kinds of indentors (glass, plexiglas, polycarbonate) were used to impose various stimuli: two different directions (forward and backward), four different stroke velocities (2, 4, 8, 16 mm/s), and eight different indentation depths (0.5 to 2.25 mm at 0.25 mm increments). Five subjects were tested with each of these stimuli. Before each session of the experiment, the fingertpad of each subject was indented eight times according to a preconditioning protocol. This preconditioning has been empirically observed to help obtain repeatable results during the subsequent indentation experiments.

The results show the relationships between normal force, shear force, and friction coefficient over time and stroke displacement. The traces of the normal force versus time consistently show an initial growth in force and then a sudden drop to an even lower level than before stroking. The curves of the shear forces grow significantly at the beginning of stroking and then drop after the peak of the related normal force curve, which is the transition point where slipping occurs. The curves of the friction coefficients versus time in all experiments grow continuously with different slopes before and after slipping of the stimulus plate. The results also enabled the determination of the relationships between stroke velocities and the friction coefficient, indentation depths and friction coefficient, and the trend of normal forces and shear forces during stroking. The friction coefficient, defined as a ratio of shear force and normal force here, is independent of stroke velocity for small indentations (around 1.0 mm). For large indentations (around 2.0 mm), when stroke velocity increases, the friction coefficient increases. Interestingly, the effect of indentation depth is the opposite: when the indentation depth increases, the coefficient decreases. This result matches the frictional behavior of rubber-like materials.

1.15.4 Identification and Control of Haptic Systems: A Computational Theory

This project provides a theoretical framework for haptics, the study of exploration and manipulation using hands. Be it human or robotic research, an understanding of the nature of contact, grasp, exploration, and manipulation is of singular importance. In human haptics, the objective is to understand the mechanics of hand actions, sensory information processing, and motor control. While robots have lagged behind their human counterparts in dexterity, recent developments in tactile sensing technology have made it possible to build sensor arrays that in some way mimic human performance. We believe that a computational theory of haptics that investigates what kind of sensory information is necessary and how it has to be processed is beneficial to both human and robotic research.

Human and robot tactile sensing can be accomplished by arrays of mechanosensors embedded in a deformable medium. When an object comes in contact with the surface of the medium, information about the shape of the surface of the medium and the force distribution on the surface is encoded in the sensor signals. The problem for the central processor is to reliably and efficiently infer the object properties and the contact state from these signals. In the first part of the project, we focused on the surface signal identification problem: the processing of sensor signals resulting in algorithms

and guidelines for sensor design that give optimal estimates of the loading and displacement distributions on the surface of the fingerpad. In the second part of the project, we focused on how the information obtained from such optimal sensing can be used for exploration of objects. We argue that an accurate reconstruction of object properties can occur using three basic building blocks of object property identification, control strategy and finger control. Object property identification pertains to the problem of inferring object properties such as shape, texture and compliance from the estimated surface signals. Control strategies define the kind of actions that are needed to identify the relevant object properties. Finally, finger control refers to the transformation of the control strategies into command trajectories, which define the desired directions of finger movement. We define and analyze the components of each of these blocks, provide explicit mathematical formulation, and give numerical examples where appropriate. Our formulation of this computational theory of haptics is independent of implementation so that it is applicable to both robots and humans.

1.15.5 Publications

Chen, J.-S. *Human Haptic Interaction with Soft Objects: Discriminability, Force Control, and Contact Visualization*. Ph.D. diss., Dept. of Mech. Eng., MIT, 1996.

