Chapter 1. Sensory Communication

Academic and Research Staff

Professor Louis D. Braida, Nathaniel I. Durlach, Dr. Charlotte M. Reed, Dr. Mandayam A. Srinivasan, Dr. Patrick M. Zurek, Dr. Cagatay Basdogan, Dr. Joseph G. Desloge, Dr. Julie E. Greenberg, Dr. Matthew H. Power, Dr. J. Kenneth Salisbury, Dr. David W. Schloerb, Dr. Barbara G. Shinn-Cunningham, Dr. Thomas E.v. Wiegand, Andrew R. Brughera, Merry A. Brantley, Lorraine A. Delhorne, Seth M. Hall, David S. Lum

Visiting Scientists and Research Affiliates

Dr. Laurel H. Carney, Dr. Steven H. Colburn, Dr. Paul Duchnowski, Dr. Kalman I. Glantz, Dr. Susan L. Goldman, Dr. Kenneth W. Grant, Dr. Jack Kotik, Dr. Karen L. Payton, Dr. William M. Rabinowitz, Dr. Christine M. Rankovic, Dr. Brigitte Schulte-Fortkamp, Dr. Kaoru Sekiyama, Dr. Melvyn Slater, Dr. Annie H. Takeuchi, Dr. Mark J. Tramo, Walter A. Aviles, Michael Boccara, Ann K. Dix, Peninah S. Fine, Geoffrey L. Plant

Graduate Students

Andrew G. Brooks, Douglas S. Brungart, Joshua P. Cysyk, Suvranu De, Matthew E. Esch, Joseph A. Frisbie, Isaac J. Graf, Chih-Hao Ho, Alexandra I. Hou, Salim F. Kassem, Kari Anne H. Kjolaas, Jean C. Krause, Jhung-Chi Liao, Samuel R. Madden, David H. Manowitz, Michael K. Qin, Balasundar I. Raju, Jason J. Sroka, Xudong Tang, Dinh-Yen T. Tran, Kimberly J. Voss, Evan Wies, Wan-Chen Wu, Hanfeng Yuan, Jonathan L. Zalesky

Undergraduate Students

Smriti Banthia, David E. DiFranco, Lajos Molnar, Ziad Obermeyer, Adrienne Slaughter, Jeremy L. Warner, Jack Yu

Technical and Support Staff

Rebecca L. Garnett, Eleanora M. Luongo, Denise M. Rossetti

1.1 Hearing Aid Research

Sponsor

National Institutes of Health
Grant R01-DC00117

Project Staff

Professor Louis D. Braida, Dr. Paul Duchnowski, Karen L. Payton, Dr. Matthew H. Power, Dr. Christine M. Rankovic, Dr. Charlotte M. Reed, Dr. Mark J. Tramo, Dr. Patrick M. Zurek, Paninah S. Fine, Isaac J. Graf, Jean C. Krause, Michael K. Qin, Jason J. Sroka, Dinh-Yen T. Tran

1.1.1 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.

1.1.2 Studies and Results

Characteristics of the Speech Signal

Clear Speech

Sentences spoken "clearly" are more intelligible than those spoken "conversationally" for hearing-impaired listeners in quiet as well as for both normal hearing and hearing-impaired listeners in noise and reverberation. Our previous work in this area has shown that the intelligibility advantage of clear speech over conversational speech is roughly 17 percentage points and has identified a number of acoustical differences between clear and conversational speech.

Since clear speech is significantly more intelligible than conversational speech for impaired listeners with mild or moderate losses in many environments, signal processing approaches for hearing aids that convert conversational speech to a close approximation of clear speech have the potential to improve speech intelligibility dramatically in many situations.

To facilitate the development of such signal processing schemes, phonetic level acoustical measurements have been made on a substantial portion (600 sentences produced by two talkers) of a phonetically labelled recorded database.

Initial measurements have focused on long-term speech characteristics, e.g., spectra, fundamental frequency distribution, distribution of pause lengths, duration and power of phones, and rate of occurrences of phonological phenomena such as vowel modification, burst elimination, alveolar flaps, and sound insertions.

The measured differences between conversational/normal and clear/slow speech are largely consistent with those of Picheny et al.[1] and Uchanski et al.,[2] with the exception of fundamental frequency (*F*0) distribution. Picheny et al. reported a wider range of *F*0 for clear/slow speech, with a slight bias towards a higher *F*0 relative to conversational/normal speech for all three (male) talkers in the study. For the two talkers examined in the current study, this difference was evident only in the male talker, whose *F*0 range increased from roughly 70 Hz for conversational/normal speech to 120 Hz for clear/slow speech. Although a similar increase in *F*0 range was not observed for the female talker, this may be due to the fact that her range of roughly 150 Hz for conversational/normal speech was already relatively large. Such an *F*0 range is typical of highly intelligible female talkers. To date, the most striking difference between conversational/normal and clear/normal speech is that the clear/normal speech contains relatively more power above 1 kHz than conversational speech.

In addition, measurements of the spectra of band amplitude envelopes indicate that envelope components with modulation frequencies below 3 Hz have greater amplitudes in clear/normal speech for the 250 Hz-2000 Hz octave bands. This finding will be examined in more detail and compared to differences in envelope spectra between conversational/normal and clear/slow speech.

Effects of Noise on Intelligibility for Tone Languages

Methods of predicting speech intelligibility under adverse listening conditions, such as the Articulation Index and the Speech Transmission Index, have generally been developed and tested for languages, such as English, in which lexical distinctions are not conveyed by changes in voice pitch. However several languages use pitch patterns, called lexical tones, to create such distinctions. Mandarin Chinese has four such lexical tones. For example, the syllable "ma", when pronounced with a falling pitch pattern means "to scold"; when pronounced with a rising pattern, the meaning is "hemp"; when pronounced with a level pattern, the meaning is "mother"; and lastly, when pronounced with a dipping pattern, the meaning is "horse". To estimate how well the intelligibility prediction methods can be used for tonal languages, we examined how well native speakers of Mandarin Chinese identify syllables differing in tones and/or articulations in the presence of additive speech spectrum noise.

The speech materials consisted of six consonant-vowel (CV) syllables and their four corresponding lexical tones, totaling 24 phonemes. A single male native speaker of Mandarin Chinese produced, in isolation, all 24 phonemes. At 0 dB SNR, the target

1 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: 434-46 (1996).

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

speech and the noise masker had the same RMS value over the duration of the target speech. Based on the results of preliminary testing, the four test conditions chosen had signal to noise ratios (SNR) of 99.9 dB, -5 dB, -10 dB, and -13 dB. The phonemes were presented to the subjects' right ears via headphones at roughly 75 dB SPL. All three listeners were native speakers of Mandarin Chinese.

Averaged across listeners, articulation and tone identification scores showed nearly the same dependence on SNR, with identification scores near perfect at 0 dB SNR and near chance at -13 dB SNR. This indicates that speech spectrum noise has the same effect on tone identification as it does on articulation identification and suggests that methods for predicting intelligibility should be applicable to tonal languages. A more fine-grained analysis indicated that the phoneme identification score was usually greater than the product of the articulation and tone identification scores, indicating that the processing of lexical tones and articulations are not independent. This implies that phoneme identification is more complex than simply the separate identification of its tone and articulation.

Computational Model of Intelligibility

Initial efforts to develop a computational model of speech intelligibility[3] were generally capable of predicting the effects of filtering on intelligibility but were not capable of predicting the effects of additive noise. The speech to noise ratio required to produce the same intelligibility as observed at the crossover for high- and low-pass filtering is +20 dB for the LCB and +10 dB for the EIH representation, as compared to roughly 0 dB for human listeners.

This excessive sensitivity to the effects of additive noise has also been observed by researchers attempting to develop automatic speech recognition (ASR) systems that are robust to the effects of interference.

To determine whether the spectral representations currently being considered for use in such systems would improve intelligibility predictions of the computational model, we compared the recognition performance that could be achieved with different spectral parameters with that measured for human listeners for /∝/-CVC syllables degraded by filtering and noise.[4] Recognition parameters included Mel-filter bank energies (MFB), Mel-frequency cepstral coefficients (MFCCs), and the ensemble interval histogram (EIH). All of the recognition systems applied a common hidden Markov model (HMM) to these parameters, so that the systems differed only in the way in which the speech signals were represented.

The recognizers were trained and tested on a database of 872 CVC syllables, including 12 consonants and 6 vowels, spoken by one male and one female. Each ASR system was trained on a subset of the materials processed by the same degradation that was applied to the (distinct) test materials. The main results of the human and ASR recognition tests were sets of confusion matrices for consonants.

Results were analyzed in terms of information transfer (IT) scores and metric multi-dimensional scaling (MMDS).[5] The IT scores measure how consistently responses were made to the phonetic features of the consonant stimuli (voicing, frication, sibilance, and place). MMDS derives a spatial configuration of the stimuli in which distances between different consonant pairs can be meaningfully compared. Both IT and MMDS measures indicated that the ASRs performed comparably in each distortion condition. For example, the average correlation between MMDS distances for different sets of recognition parameters was 0.92. Unlike studies of consonant recognition by humans that used white (as opposed to speech-spectrum) noise, highpass and additive noise degradations produced similar confusion matrices that differed from the confusion matrices for lowpass filtering.

When the syllables were degraded by highpass or lowpass filtering, the recognition scores for the ASR systems were comparable to those obtained by human listeners. Additive noise, however, reduced scores for the ASR systems to a much greater extent than those for human listeners. To achieve similar scores in noise, the ASR systems required roughly a 10 dB higher signal to noise ratio than human listeners. MMDS analyses indicated that for human listen-

3 M.H. Power and L.D. Braida, “Consistency Among Speech Parameter Vectors: Application to Predicting Speech Intelligibility,” *J. Acoust. Soc. Am.* 100: 3882-98 (1996).

4 J.J. Sroka, *Consonant Recognition by Humans and Machines*, Ph.D. thesis, Division of Health Sciences and Technology, MIT, September 1998.

5 L.D. Braida, “Crossmodal Integration in the Identification of Consonant Segments,” *Quart. J. Exper. Psych*. 43A: 647-77 (1991).

ers the configurations for consonants under lowpass filtering and additive noise degradations were similar (a correlation of 0.76). For the ASR systems, however, the configurations for the additive noise degradations were more similar to those for highpass filtering than for lowpass filtering (a correlation of 0.84 versus 0.58). MMDS comparisons between human listeners and ASR systems showed greater similarity for the noise degradation than for lowpass or highpass filtering (correlations of 0.84, 0.50, and 0.50, respectively). Thus, while error rates for the ASR systems were fairly similar to those for human listeners for the filtering conditions, the patterns of confusions were not. Conversely, the ASR systems achieved lower recognition scores than human listeners for speech degraded by additive noise, but the patterns of confusions were similar.

The discrepancy between the consonant recognition performance in noise of humans and ASR systems is greatest for the phonetic feature *voicing*. ASR systems require a 10 dB more favorable SNR to make voicing distinctions as well as humans. Analysis of the waveforms used in the human consonant identification tests revealed two robust cues to the voicing distinction that may not have been used efficiently by the ASR systems: voice onset time (VOT), and change in fundamental frequency ($F0$) at the start of the vowel. Incorporating automatically made measures of these cues with traditional ASR parameters improved voicing scores by 5-7 percentage points for the ASRs, roughly half of the difference between human and ASR scores at 0 dB SNR.

Integration Models of Intelligibility

Part of our research on models of speech intelligibility is addressed to the question of how well cues are integrated across frequency bands. Although listeners with normal hearing typically exhibit near-perfect integration, patients with cortical lesions may exhibit deficiencies on integration tasks. But such deficiencies can be obscured if the representation of spectral variables is altered. Microelectrode studies in nonhuman primates and other mammals have demonstrated that many neurons in auditory cortex are excited by pure tone stimulation only when the tone's frequency lies within a narrow range of the audible spectrum. However, the effects of auditory cortex lesions in animals and humans have been interpreted as evidence against the notion that neuronal frequency selectivity is functionally relevant to frequency discrimination.

Here we report psychophysical and anatomical evidence in favor of the hypothesis that finegrained frequency resolution at the perceptual level relies on neuronal frequency selectivity in the cortex. An adaptive procedure was used to measure frequency difference thresholds for frequency discrimination in five humans with focal brain lesions and eight normal controls. Only the patient with bilateral lesions of superior temporal cortex that included all of AI and much of the surrounding auditory cortex showed markedly elevated frequency difference thresholds. Weber fractions for relative pitch discrimination were about eight-fold higher than the Weber fractions associated with unilateral lesions of auditory cortex, auditory midbrain, or frontal cortex; Weber fractions for same-different discriminations were about seven times higher. In contrast, intensity thresholds for pure tone detection, difference thresholds for pure tone duration discrimination at long center durations, difference thresholds for vibrotactile intensity discrimination, and judgments of visual line orientation were normal or only mildly impaired following bilateral lesions of AI and surrounding cortex. We interpret the present results as evidence that: fine-grained frequency resolution at the perceptual level relies on the integrity of finely-tuned neurons in AI and surrounding fields.

Journal Article

Tramo, M.J., G.D. Shah, and L.D. Braida. “The Functional Role of Auditory Cortex in Frequency Processing.” *J. Neurophysiol*. Under revision.

Thesis

Sroka, J.J. *Consonant Recognition by Humans and Machines.* Ph.D. thesis. Division of Health Sciences and Technology, MIT, September 1998.

1.2 Enhanced Communication for Speechreaders

Sponsor

National Institutes of Health
Grant R01 DC02032

Project Staff

Professor Louis D. Braida, Dr. Paul Duchnowski, Dr. Susan L. Goldman, Dr. Matthew H. Power, Dr. Annie H. Takeuchi, Lorraine A. Delhorne, Dr. Charlotte M. Reed, Dr. Kaoru Sekiyama, Dr. Kenneth W. Grant, Smriti Bamthia, Ann K. Dix, Joseph A. Frisbie, Jean C. Krause, David S. Lum, Ziad Obermeyer, Jeremy L. Warner, Jack Yu

1.2.1 Specific Aims

Our long-term goal is to develop aids for individuals with hearing impairments so severe that their communication relies heavily on speechreading. 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 can improve speech reception markedly when the cues they provide are integrated with cues derived from the visible actions of the talker's face. 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.

Models of Audiovisual Integration

According to the pre-labeling model of audiovisual integration[6] audiovisual (AV) stimuli are identified on the basis of the sample value of a vector of cues that has two sets of components, auditory and visual, whose properties are assumed to be independent. Although the model has been applied to the results of AV studies that used only compatible stimuli in previous work, we have begun to apply it to AV tests that included both compatible (e.g., auditory /ba/ paired with visual /ba/) and incompatible (e.g., auditory /ba/ paired with visual /ga/) stimuli.

In the tests, 28 young Japanese adults were asked to specify the stimulus from the set /ba, da, ga, pa, ta, ka, ma, na/. The syllables were spoken in isolation by two native speakers of Japanese and two native speakers of American English. Incompatible audiovisual stimuli (with discrepant place of articulation) were constructed by dubbing. To facilitate analysis of the confusion patterns in the unimodal conditions, we used lowpass filtering and additive noise to degrade the auditory stimuli and diffusion screens to degrade the visual stimuli.

In the model, stimuli are characterized by stimulus centers whose locations can be estimated by separate analyses of the auditory and visual confusion matrices obtained in experiments that employ unimodal presentation conditions. Responses are determined by the proximity of the cue vector to response centers. In AV experiments that employ exclusively compatible stimuli, the response centers are typically close to the corresponding stimulus centers. According to the model, the same analysis of unimodal experiments serves to specify the stimulus centers for incompatible AV stimuli. We have determined that, at least under some conditions, the responses to these stimuli correspond to the use of the same response centers as for the compatible stimuli.

Confusion matrices were predicted for the bimodal stimuli directly from the stimulus and response centers derived from the matrices for the unimodal experiments without modification and with no free parameters introduced. The patterns of responses to both the compatible and incompatible stimuli were reasonably well predicted by the pre-labeling model. This suggests that the occurrence of the McGurk effect is a manifestation of the same integration processes evident in “ordinary” audiovisual speech reception for compatible stimuli. Moreover, the magnitude of the effect can be predicted from measurements made under unimodal presentation conditions alone. For the stimuli considered, this magnitude seems to be determined by the relative distinctiveness of the auditory and visual stimuli. In particular,

6 L.D. Braida, “Crossmodal Integration in the Identification of Consonant Segments,” *Quart. J. Exper. Psych*. 43A: 647-77 (1991).

the labial stimuli were more distinct from the non-labial stimuli in the visual presentation condition than in the auditory presentation condition. As a result, when labial visual stimuli are paired with non-labial auditory stimuli, the resulting incompatible bimodal stimuli are more similar to the compatible labial stimuli than to compatible non-labial stimuli.

Meeting Paper

Braida, L.D., K. Sekiyama, and A.K. Dix. "Integration of Audiovisually Compatible and Incompatible Consonants in Identification Experiments." Paper published in the *Proceedings of the Audio-Visual Speech Processing Conference* (AVSP98), New South Wales, Australia, December 4-6, 1998.

Supplements Based on Signal Processing

Speechreading 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. Amplitude envelopes can be derived by bandpass filtering the speech signal, full-wave rectifying the filtered signal, and smoothing the rectified signal via lowpass filtering. The resulting envelope signals are conveyed to the listener by amplitude modulating one or more tones. Mansour[7] has shown that a pair of envelope signal (derived from the octave-bands of speech centered at 500 and 3000 Hz) and conveyed by modulating the amplitudes of 200 and 240 Hz can enhance speechreading substantially. To explore this finding more thoroughly, we conducted two additional experiments to determine the effects of a variety of such signals on the intelligibility of words in low-context sentences and on the identification of consonants in nonsense syllables.

One experiment evaluated the usefulness of band-envelope signals as supplements to speechreading for low-context sentences. The test conditions included a subset of the conditions studied previously: (a) speechreading alone and speechreading supplemented by (b) a 200 Hz tone whose amplitude was modulated by the envelope of the 500 Hz octave band of speech, (c) a 200 Hz tone whose amplitude was modulated by the envelope of the 500 Hz octave band of speech plus a 3000 Hz tone whose amplitude was modulated by the envelope of the 3000 Hz octave band of speech, (d) a 200 Hz tone whose amplitude was modulated by the envelope of the 500 Hz octave band of speech plus a 500 Hz tone whose amplitude was modulated by the envelope of the 3000 Hz octave band of speech, (e) a 200 Hz tone whose amplitude was modulated by the envelope of the 500 Hz octave band of speech plus a 240 Hz tone whose amplitude was modulated by the envelope of the 3000 Hz octave band of speech, and (f) speech that had been lowpass filtered to 240 Hz. All stimuli were presented binaurally via free field. Results indicated that presentation of band envelope signals improved keyword scores relative to speechreading alone (29%) with the highest scores (77%) obtained with supplement (c). The scores (58-64%) obtained with supplements (b) and (d,e,f) were roughly equal. These results differ substantially from those obtained by Mansour,[7] who obtained similar scores (76 and 74%) with supplements (c) and (e) that were higher than that (63%) obtained with supplement (b). Further studies are planned to investigate the cause of this discrepancy.

A second experiment attempted to evaluate the usefulness of supplements (b), (c), and (e) to the speechreading of initial consonants, from a set of 24, in isolated CV syllables. Identification scores were 43% by speechreading alone, and 15, 21, and 16% by the auditory supplements alone, respectively. When combined with speechreading, the supplements improved identification scores to 56, 64, and 58%, respectively. These results are consistent with those obtained in the first experiment in that supplement (c) was more effective than supplements (b) and (e), which were equally effective. These results will be analyzed further to determine differences in the pattern of consonant confusions obtained with the three supplements.

Supplements Based on Automatic Speech Recognition

Manual cued speech is a system of hand gestures designed to help deaf speechreaders distinguish among ambiguous speech elements. We have developed a computerized cueing system that uses automatic speech recognition to determine and display cues to the cue receiver.[8] Keyword scores of 66% in

7 S.A. Mansour, *Effects of Amplitude Envelope Cues as an Aid to Enhanced Speechreading,* M.Eng. thesis, Department of Electrical Engineering and Computer Science, MIT, May 1997.

8 P. Duchnowski, L. Braida, M. Bratakos, D. Lum, M. Sexton, and J. Krause (1998), "A Speechreading Aid Based on Phonetic ASR," in Proc. Int. Conf. Spoken Lang., December 1998.

low-context sentences have been obtained with this system, nearly double the speechreading-alone scores. Although improving cue recognition accuracy is expected to improve the effectiveness of the automatic cuer, it appears that substantial gains can also be obtained by improving the characteristics of cue display. In particular, gains are likely to result from increasing the distinctiveness of the cue shapes and from altering the temporal alignment of the cues with the visible actions of the face of the talker.

Even highly experienced MCS receivers make errors on perfectly-cued low-context sentences spoken at 100 wpm, suggesting that 10-20% of the segments are perceived incorrectly. Preliminary analysis of error responses suggests that more than one-quarter of the word errors can be attributed to incorrect reception of cues for segments in words. One way to reduce the number of such errors may be to alter the appearance of the surface or the outline of the hand. Two easily confused shapes could be displayed differing in brightness, size, or color. In intial work, we focused on cue coloration to improve discrimination.

The experiments tested the discriminability of hand shapes before and after some of the hand images were digitally colored.[9] We used pre-recorded C1-/ah/-C2 syllables spoken by one female talker. C1 was chosen from the set of 24 initial consonants, while C2 was drawn at random from a set of eight and played no role in the experiments. The subjects were normal-hearing adults, none of whom had any familiarity with MCS.

In one experiment, subjects identified both the initial consonant and the hand shape that was superimposed on the CVC recording for 33 msec near the beginning of the syllable. Tests were conducted both with naturally colored handshapes and with a set of hand images consisting of three colored (blue, red, and green) and five untreated images. Consonant identification scores (roughly 44%) were unaffected by the introduction of coloration. Handshape identification scores, on the other hand, increased by roughly seven percentage points, from 86 to 93% correct, when colorization was introduced.

In a second experiment, subjects identified the second hand shape, in a sequence of three, that was superimposed on the CVC images. The duration of target shape was 200 msec, while the surrounding shapes had durations of 66 msec each. In addition to the three hand images of the first experiment, two additional images were colorized (yellow and violet). Results indicated that scores for consonants were roughly the same as in the first experiment and were unaffected by colorization. Hand shape identification scores, however, improved by roughly nine percentage points, from 68 to 77%, when colorization was used.

Taken together, these experiments suggest that selective colorization of hand images may improve cue reception. Although the improvement in shape recognition accuracy does not appear to be large, it can be achieved by only minor modification to the automatic cueing system. Since nearly all cue reception errors lead to errors in word recognition when context is limited, such modifications may be highly beneficial. Tests of the reception of cued speech using colorized hand images are currently being planned.

The results of our preliminary studies of automatic cueing systems suggest that temporal misalignment between the visible actions of the face during speech production and the presentation of synthetic cues can have a large effect on the usefulness of the cues. We are exploring this effect more systematically by evaluating the intelligibility of cued speech as a function of this delay. As a reference condition, we are also evaluating the intelligibility of speech degraded by noise when supplemented by speechreading. In initial tests, subjects who obtained an intelligibility score of 18% on low-context materials in the speechreading alone (SA) condition, obtained scores of 85% when the noisy speech was presented simultaneously with visual facial actions. The effect of temporal misalignment appeared to be asymmetric, with a given delay in the presentation of the visual signal more deleterious than in the presentation of the audio signal. Thus, for example, delaying the audio signal by 100 ms had little effect on intelligibility, while advancing the audio signal by 300 ms produced scores that were similar to those in the SA condition. Parallel experiments on the reception of cued speech are planned.

Publications

Duchnowski, P., L. Braida, M. Bratakos, D. Lum, M. Sexton, and J. Krause. “A Speechreading Aid

9 P. Duchnowski, L. Braida, D. Lum, M. Sexton, J. Krause, and S. Banthia, “Automatic Generation of Cued Speech for the Deaf: Status and Outlook,” *Proc. Audio-Visual Speech Proc. Conf.* (AVSP’98), September 1998.

Based on Phonetic ASR." *Proc. Int. Conf. Spoken Lang*. December 1998.

Duchnowski, P., L. Braida, D. Lum, M. Sexton, J. Krause, and S. Banthia. "Automatic Generation of Cued Speech for the Deaf: Status and Outlook." *Proc. Audio-Visual Speech Proc. Conf*. (AVSP'98). September 1998.

1.3 Tactile Communication of Speech

Sponsors

National Institutes of Health/National Institute on Deafness and Other Communication Disorders
Grant 2 R01 DC00126

Project Staff

Andrew R. Brughera, Lorraine A. Delhorne, Nathaniel I. Durlach, Seth M. Hall, Eleanora Luongo, Geoffrey L. Plant, Charlotte M. Reed, Mandayam A. Srinivasan, Hanfeng Yuan

1.3.1 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.

Research during the past year was conducted in the four areas described below.

1.3.2 Basic Studies of Hand Stimulation and Active Touch

A series of studies has been conducted to explore the capabilities of the tactual sense for conveying information. These studies employ a multi-finger tactual display designed by Tan[10] to provide stimulation along a continuum from high-frequency, low-amplitude kinesthetic movements to low-amplitude, high-frequency cutaneous vibrations. Current experiments are being conducted to explore the effects of masker characteristics on the identification of particular types of targets.[11] This study employs a set of seven signals at two different durations (125 or 250 msec) presented to the index finger of the left hand. Subjects' ability to identify these signals is being tested with three different masking paradigms: forward, backward, and "sandwich" masking. In the forward and backward masking paradigms, subjects are presented with two signals (separated by six values of inter-stimulus interval in the range of 0 to 640 msec) and asked to identify either the second signal (in the forward-masking conditions) or the first signal (in the backward-masking conditions). In the "sandwich" masking paradigm, subjects are presented with three successive stimuli (again using inter-stimulus intervals in the range of 0 to 640 msec) and asked to identify the middle signal.

The stimulus set was selected to permit examination of the temporal-integration properties of the maskers and targets under each of the paradigms. The stimuli at each of the two durations consisted of (1) three single frequency waveforms (low, middle, and high frequency), (2) three double-frequency waveforms (resulting from all possible combinations of the three single waveforms), and (3) one triple-frequency waveform (resulting from the combination of the three single frequency waveforms). Error responses were compared to the target to determine the extent to which errors arose from the use of the masker as a response, the combination of the masker and the target, and the combination of a single component of a multicomponent masker with the target. The results indicate that, across stimulus duration and ISI, roughly 20-35% of all errors arise from the use of the masker as the response; there is no evidence of the use of a composite response formed by combining spectral components across the masker and the target; and while there is some evidence of the intrusion of mid- and high-frequency masker components into the response, this does not appear to occur for low-frequency components.

10 H.Z. Tan, "Information Transmission with a Multi-Finger Tactual Display," Ph.D. diss. Department of Electrical Engineering and Computer Science, MIT, 1996.

11 H.Z. Tan, N.F. Wan, and C.M. Reed, "Masking and Temporal Integration with a Multi-Finger Tactual Display," paper presented at International Sensory Aids Conference, Portland, Maine, May 11-13, 1998.

1.3.3 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. Work has continued on the development of speech communication skills with a small group of children (at the Randolph Learning Center) who have been transferred from the Tactaid 7 device to the Tactilator and/or the Tactaid 2000+ (a display of high frequency speech energy in the range of 2-7 kHz). In one case, work is being conducted with one child with good low-frequency hearing but little hearing in the higher frequencies using the Tactaid 2000+ in conjunction with the child's hearing aids. Work has been conducted on the development of test procedures to evaluate the ability of deaf children to detect, discriminate, and identify the sibilant and fricative consonants.

A two- and three-alternative forced-choice test suitable for use with elementary school-age children has been designed and will be recorded by male, female, and child speakers. This test will be administered to a group of hard-of-hearing and profoundly deaf children attending the Rhode Island School for the Deaf. A five-alternative forced-choice test contrasting the 22 initial consonants of English in meaningful words has been designed and will be recorded on videotape for use in the evaluation of tactile aids as supplements to lipreading. An analysis of the syllabic structure and the occurrence of consonants in the 500 most common words of spoken American English (Dahl's 1979 corpus) has been conducted for application to the development of more appropriate testing and training procedures including sets of sentences and word lists.

1.3.4 Development of Improved Tactual Supplements to Speechreading

Work during the past year has focused on implementing the multifinger tactual display developed by Tan[12] for displaying speech signals. Software and hardware modifications have been implemented such that the device is now capable of conveying amplitude-envelope signals that are derived from filtered bands of speech. In addition, software is being developed to control experiments employing audio-visual speech materials which are recorded on laser discs. Thus, a system will soon be in place for creating multidimensional signals derived from speech for presentation on our multifinger tactual display and for evaluating the contributions of these signals as an aid to speechreading.

1.3.5 Study of the Reception of Environmental Sounds through Tactual Displays

Responses from self-assessment questionnaires completed by profoundly deaf adult users of tactile aids indicate that these individuals place a high value on the information provided by these devices for receiving and interpreting non-speech sounds in the environment. Although tactile aids have been designed primarily to convey aspects of the acoustic speech signal (and have typically been evaluated as such), they obviously are also a source of information about other acoustic signals in the environment. In order to develop a more complete evaluation of the benefits provided by tactile aids, we have developed a test of the reception of environmental sounds and compared results from experienced tactile-aid users to those of laboratory-trained subjects and subjects with cochlear implants.[13]

Development of a test of environmental-sound reception. To provide an assessment of environmental-sound reception under conditions that approximate those encountered in natural situations, a test was developed employing sets of ten sounds likely to occur in each of four different settings (office, outdoors, kitchen, and general home environments). The stimuli for each of the tests included environmental sounds from the following categories: water-produced sounds (e.g., water bubbler and running brook), discrete impacts (e.g., door slam and footsteps), modulated noise (e.g., vacuum cleaner and car engine), multiple mechanical transients (e.g., stapler and keyboard), warning signals (e.g., telephone and siren), and a final class of animal, human, and music sounds (e.g., dog bark, baby cry, and orchestra). The test stimuli were obtained from a CD-ROM recording of sounds for use in special effects in the movie industry. Three tokens of each sound were

12 H.Z. Tan, "Information Transmission with a Multi-Finger Tactual Display," Ph.D. diss., Department of Electrical Engineering and Computer Science, MIT, 1996.

13 C.M. Reed, and L.A. Delhorne, "Identification of Environmental Sounds through Tactile Aids," paper presented at International Sensory Aids Conference, Portland, Maine, May 11-13, 1998; L.A. Delhorne, and C.M. Reed, "Identification of Environmental Sounds through Cochlear Implants," paper presented at the International Sensory Aids Conference, Portland, Maine, May 11-13, 1998.

created, either by sampling different waveforms from the CD-ROM (e.g., three different bird songs) or by sampling three separate sections of a long waveform (e.g., three separate 2-sec samples of thunder from a 30-sec recording). The waveforms, which were sampled and stored at a 22 kHz rate, varied in duration (from roughly 0.5 to 5 sec) and in amplitude (± 8 dB from mean RMS) across sounds. Software was developed in MATLAB for presenting stimuli and recording responses using a one-interval, 10-alternative, forced-choice procedure either with or without correct-answer feedback. On each trial, the ten alternatives appeared on a computer terminal, one of the 30 waveforms was selected at random without replacement, the waveform was played out using direct input to the tactile device, and the subject used a computer mouse to click on his/her response. Data files that kept a record of the stimulus and response from each trial were then used to generate stimulus-response confusion matrices from which measures of percent-correct and information transfer were computed.

Results with laboratory-trained subjects. Data were obtained from two normal-hearing laboratory subjects with the Tactaid 7 device (with direct input) worn on the forearm and earplugs and earmuffs to eliminate any auditory cues. Signals were presented at a comfortable level of approximately 25 dB SL. For each of the four settings (office, outside, kitchen, general home), subjects received 600 trials with correct-answer feedback followed by an additional 300 trials without feedback. Post-training performance, across settings and subjects, averaged 57% correct. Performance improved by roughly 20 percentage points over the course of the training. A third subject received extensive training with both the Tactaid 7 and Tactaid 2 devices. Post-training scores indicated that performance was similar for the two devices, averaging 79% for the Tactaid 7 and 76% for the Tactaid 2. Thus, there does not appear to be a clear advantage for the more detailed spectral information provided by the Tactaid 7 in identifying small sets of environmental sounds.

Results with deaf adult users of tactile aids. Three profoundly deaf subjects (regular users of the Tactaid 7 device who participated in the field study of tactile aids reported on previously[14]) were tested on their ability to recognize environmental sounds. Within each of the four environments, subjects were tested initially on 300 trials without correct-answer feedback to determine their ability to identify the sounds based solely on their own prior experience with their tactile aids. The subjects then received an additional 300 trials with the presentation of trial-by-trial correct-answer feedback to determine the effects of training on performance of the task. Individual-subject results are summarized in Table 1 below for each test environment, with and without the use of correct-answer feedback.

Table 1: Summary of results for each test environment.

Subject	Kitchen		General Home		Office		Outdoors	
	Feedback		Feedback		Feedback		Feedback	
	No	Yes	No	Yes	No	Yes	No	Yes
JL	49.3	68.7	41.3	61.3	50.0	70.0	43.0	61.7
MC	51.4	63.9	36.2	72.0	44.3	77.2	36.2	63.3
RS	14.0	45.0	18.7	42.0	30.0	47.0	19.3	33.7
Average	38.2	59.2	32.1	58.4	41.4	64.7	32.8	52.9

The results indicate that the performance of subjects JL and MC was similar and superior to that of subject RS; performance, which was similar across the four environments, averaged 38.1% correct for initial testing without feedback; and scores were increased by an average of 21 percentage points in the presence of correct-answer feedback. Thus, while feedback led to substantial improvements in performance, subjects were able to identify closed sets of environmental sounds at levels greater than that expected on the basis of chance alone even without any specific training on the task. These results imply that cues useful to the identification of non-speech environmental sounds are provided by the seven-channel formant-

14 C.M. Reed, and L.A. Delhorne, "Current Results of a Field Study of Adult Users of Tactile Aids," *Seminars in Hearing* 16: 305-15 (1995).

based display of the Tactaid 7 device. Additional results obtained with subject MC under conditions of hearing aid alone and hearing aid combined with tactile aid indicate that her scores for the combined condition were substantially higher than those obtained with either the hearing aid or tactile aid alone. These results suggest that the information derived from the tactile aid was to some extent independent of that derived from the hearing aid and that the information from the two sources was integrated.

Results with adult users of cochlear implants. The environmental-sound reception test was administered to a group of individuals implanted with either the Ineraid or Clarion device who had participated in long-term evaluations of their speech-reception ability.[15] Data were collected initially without correct-answer feedback; however, depending on level of performance, additional data were collected for some subjects and environmental settings using trial-by-trial correct-answer feedback. Of the eleven subjects, performance across settings averaged 90% or greater for 5 subjects; 70-80% correct for four subjects; 60-70% for 1 subject; and less than 50% for one subject. Ability on this task appears to correlate well with the ability to identify monosyllabic words through the implant alone.

1.3.6 Publications

Journal Articles

Fischer, S.D., L.A. Delhorne, and C.M. Reed. "Effects of Rate of Presentation on the Reception of American Sign Language." *J. Speech Hear. Res*. Forthcoming.

Plant, G. "Training in the Use of a Tactile Supplement to Lipreading: A Long-Term Case Study." *Ear Hear.* 19: 394-406 (1998).

Plant, G. "Speech Training for Congenitally Deaf Young Adults: A Case Study." *Volta Rev*. Forthcoming.

Reed, C.M., and N.I. Durlach. "Note on Information Transfer Rates in Human Communication." *Presence: Teleop. Virt. Reality* 7: 509-18 (1998).

Tan, H.Z., N.I. Durlach, C.M. Reed, and W.M. Rabinowitz. "Information Transmission with a Multi-Finger Tactual Display." *Percep. Psychophy.* Forthcoming.

Book

Plant, G. *Step by Step: The Foundations of Intelligible Speech.* Somerville, Massachusetts: Hearing Rehabilitation Foundation, 1998.

1.4 Multimicrophone Hearing Aids

Sponsor

National Institutes of Health
Grant 2 R01 DC00270

Project Staff

Joseph G. Desloge, Dr. Julie E. Greenberg, Dr. William M. Rabinowitz, Dr. Patrick M. Zurek

The goal of this research is to improve 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.

A new beamforming algorithm, called the location-estimating, null-steering (LENS) algorithm, has been developed and evaluated.[16] LENS processing is innovative because it uses a novel robustness-control mechanism to yield a beamformer that avoids target cancellation under adverse conditions. Most traditional beamforming systems realize robustness control through the use of constraints in the beamforming optimization, but this approach is both indirect and difficult to understand. LENS, on the other hand, achieves direct and obvious robustness control by separating robustness control from the beamforming optimization in the following two-step procedure: (1) it solves a minimally-constrained beamforming optimization in terms of the LENS parameter set B, and then (2) it evaluates and restricts any components of B that might cause target cancellation. By using this separated robustness-control approach, simulations have demonstrated that LENS can prevent target cancellation to a higher degree than several traditional beamformers.

Furthermore, the advantages of LENS processing are not limited to improved system robustness. Its design allows implementation using a relaxation-

15 W.M. Rabinowitz, D.K. Eddington, L.A. Delhorne, and P.A. Cuneo, "Relations among Different Measures of Speech Reception in Subjects using a Cochlear Implant," *J. Acoust. Soc. Am.* 92: 1869-81 (1992).

16 J.G. Desloge, *The Location-Estimating, Null-Steering (LENS) Algorithm for Adaptive Microphone-array Processing,* Ph.D. diss., Department of Electrical Engineering and Computer Science, MIT, 1998.

based approximation to direct-solution LENS processing (the LENS equivalent of direct covariance matrix inversion processing for traditional systems). Simulations have demonstrated that this relaxation-based implementation can combine efficient implementation, fast beamformer adaptation, and good beamforming performance, which is difficult to achieve with traditional systems.

Other work has dealt with the problems encountered when the traditional LMS algorithm is used in the presence of strong desired signals. A major disadvantage of the LMS algorithm is its excess mean-squared error, or misadjustment, which increases linearly with the desired signal power. This leads to degrading performance when the desired signal exhibits large power fluctuations and becomes a serious problem in many speech processing applications, including microphone-array hearing aids.

Recent work considers two modified LMS algorithms designed to solve this problem by reducing the size of the steps in the weight update equation when the desired signal is strong. Analysis of the modified LMS algorithms indicates that either one provides substantial improvements in the presence of strong desired signals and similar performance in the presence of weak desired signals, relative to the unmodified LMS algorithm. Computer simulations with both uncorrelated Gaussian noise and speech signals confirm the results of the analysis and demonstrate the effectiveness of the modified algorithms. The modified LMS algorithms are particularly suited for signals (such as speech) that exhibit large fluctuations in short-time power levels.

1.4.1 Publications

Journal Article

Greenberg, J.E. “Modified LMS Algorithms for Speech Processing with an Adaptive Noise Canceller.” *IEEE Trans. Speech Audio Process*. 6: 338-51 (1998).

Doctoral Dissertation

Desloge, J.G. *The Location-Estimating, Null-Steering (LENS) Algorithm for Adaptive Microphone-array Processing.* Ph.D. diss. Department of Electrical Engineering and Computer Science, MIT, 1998.

1.5 Hearing-Aid Device Development

Sponsor

National Institutes of Health
Contract N01 DC-5-2107

Project Staff

Merry A. Brantley, Andrew R. Brughera, Dr. Julie E. Greenberg, Dr. William M. Rabinowitz, Dr. Patrick M. Zurek

The overall objective of work under this contract is to evaluate promising signal processing algorithms for hearing aids under realistic conditions.

Our recent work assesses improvements to speech reception provided by array processing of signals from two ear-level microphones when a single interfering noise is present.[17] In our first study, we implemented three experimental algorithms and a (no-processing) reference condition. For all three of the algorithms, array processing was confined to the frequency region above 1 kHz; the signal spectrum below 1 kHz was left unmodified. This lowpass/highpass strategy was an attempt to preserve a sense of auditory space for the hearing aid user (based on use of low-frequency binaural cues) while enhancing speech reception in the high-frequency region.

Algorithms were implemented in realtime and evaluated with tests of speech reception in noise. The improvements in SRTs with the experimental algorithms over the reference condition were small. Although statistically significant, the improvement provided by an adaptive beamforming algorithm—one that we expected to perform very well in this test—was only 1.5 dB. Measurement of the interference spectra at the outputs of the hearing aids showed that the array processing algorithms were effective at reducing the interference by about 20 dB above 1 kHz. Measurements of the target spectra also demonstrated that the target speech signal passes uncorrupted through the algorithm.

In a subsequent study we eliminated the lowpass/highpass processing scheme and allowed array processing across the spectrum. With this change,

17 P.M. Zurek, J.E. Greenberg, M. Brantley, A.R. Brughera, and W.M. Rabinowitz, “Laboratory Assessment of Advanced Signal Processing Algorithms for Hearing Aids,” paper presented at the Fifth Biennial Conference on Issues in Advanced Hearing Aid Research, Lake Arrowhead, California, 1998.

speech-reception improvements in noise from adaptive beamforming for eight hearing-aid users averaged 9.8 dB.

These findings indicate that hearing-impaired listeners are unable to take advantage of large improvements in speech-to-noise ratio in the high-frequency region. This inability to make use of high-frequency speech cues combined with the restriction of array processing to the region above 1 kHz resulted in small SRT benefits for algorithms using the lowpass/highpass scheme. In light of this limitation on effective operating bandwidth of the impaired ear, we have to abandon our attempt to provide both improved speech reception and spatial perception simultaneously. In practical terms, this means that a hearing aid incorporating this type of array processing will need a switch to change between a focused directional mode and a wide-span spatial listening mode.

1.5.1 Meeting Papers

Zurek, P.M., J.E. Greenberg, M. Brantley, A.R. Brughera, and W.M. Rabinowitz. "Laboratory Assessment of Advanced Signal Processing Algorithms for Hearing Aids." Paper presented at the Fifth Biennial Conference on Issues in Advanced Hearing Aid Research, Lake Arrowhead, California, 1998.

1.6 Virtual Environment Technology for Training

Sponsor

U.S. Navy - Office of Naval Research
Grant N61339-96-K-0002
Grant N61339-96-K-0003
Grant N00014-97-1-0655

Project Staff

Nathaniel I. Durlach, Dr. Mandayam A. Srinivasan, Dr. J. Kenneth Salisbury, Dr. Thomas E.v. Wiegand, Lorraine A. Delhorne, Dr. Cagatay Basdogan, Dr. David W. Schloerb, Andrew G. Brooks, David E. DiFranco, Chih-Hao Ho, Alexandra I. Hou, Kari Anne H. Kjolaas, Samuel R. Madden, Lajos Molnar, Adrienne Slaughter, Evan Wies, Wan-Chen Wu, Hanfeng Yuan

1.6.1 Introduction

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. At RLE, two components of this program are being pursued: (1) enabling research on the human operator (ERHO) and (2) development of haptic interfaces and multimodal virtual environments. The ERHO component is concerned with how human perception and performance in virtual environments (VEs) depend upon (1) 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 evaluation of VE training systems, but also for VE systems in general. The second component is focused on the development of haptic interfaces that enable the user to touch, feel and manipulate objects in VEs. Software is being developed to generate haptic stimuli and to integrate visual, auditory, and haptic displays. Experiments on multimodal illusions due to interactions between haptic and visual or auditory displays have also been conducted. The progress over the past year in ERHO, haptic interface development, and multimodal VEs is described in the following subsections.

1.6.2 Visual Depth Perception in Virtual Environments

During the past year, we have prepared some previous results[18] related to depth perception via stereopsis for publication[19] and also conducted further work on depth perception via motion parallax.[20]

In the motion parallax study, attempts were made to determine the effects of time delay between head movement and change in the visual image presented on a desktop monitor. Computer generated random

18 D.W. Schloerb, *Adaptation of Perceived Depth Related to Changes of the Effective Interpupillary Distance in Computer-Generated Stereoscopic Displays*, Ph.D. diss. Department of Mechanical Engineering, MIT, 1997.

19 T.E.v. Wiegand, D.W. Schloerb, and W.L. Sachtler, "Virtual Workbench Near-Field Virtual Environment System with Applications," *Presence: Teleop. Virt. Environ.*, forthcoming.

20 H.F. Yuan, *Effect of Delays on Depth Perception by Motion Parallax in Virtual Environments,* M.S. thesis, Department of Electrical Engineering and Computer Science, MIT, 1998.

dot patterns were structured and related to head movement in such a manner that a sinusoidal function in depth appeared when the head was moved laterally. The phase of the sinusoid was varied randomly between 0° and 180°. In one case, the top portion of the image appeared in front of the screen and the bottom portion behind; in the other case, the opposite occurred. The task of the subject (using a two-alternative forced-choice paradigm) was to judge which of the two cases occurred. In each run, both the amplitude A of the sinusoid and the time delay T was randomly roved over a wide range ($0.01 \leq A \leq 1.00$ inches; $55 \leq T \leq 1760$ msec). The results of these experiments showed that the delay had no effect on performance for $T \leq 220$ msec. Although this delay is sufficiently large to ensure that time delay will not interfere with depth perception by motion parallax for many VE systems, this delay is likely to be exceeded when the VE system is very complex and/or involves communication over long distances.

Journal Article

Wiegand, T.E.v., D.W. Schloerb, and W.L. Sachtler. "Virtual Workbench Near-Field Virtual Environment System with Applications." *Presence: Teleop. Virt. Environ.* Forthcoming.

Thesis

Yuan, H.F. *Effect of Delays on Depth Perception by Motion Parallax in Virtual Environments*. M.S. thesis. Department of Electrical Engineering and Computer Science, MIT, 1998.

1.6.3 Part-Task Trainer for Position-Velocity Transformations

Novice ship handlers often have a problem learning to transform position-velocity information from one coordinate system to another. During the past year, we have initiated development of a part-task trainer to enhance the ability to perform transformations involving relative motion. Although, abstractly, these transformations are trivial in the sense that they only involve addition or subtraction of vectors, for many trainees, performing such transformations quickly and easily is a skill that has to be learned.

In the attempt to enhance such learning, we have begun to develop a set of desktop video exercises involving transformations among three vectors: (1) the velocity vector of the ship the trainee is piloting, (2) the velocity vector of another, independent ship, and (3) the velocity vector showing the relative motion of the two ships. In these exercises, the user is able to control one or more of these vectors with joysticks. In practice mode, the user sets two of the vectors with the joysticks, and the system computes and displays the third vector. In the test mode, the system presents two of the vectors, and the trainee is required to create the third using the joystick. Once the trainee has entered his or her vector estimate, the system provides immediate correct-answer feedback by displaying the correct vector. The system cumulates errors in both magnitude and angle, as well as in vector difference. Also recorded is the trainee's response time. Training experiments using this system are now being initiated.

1.6.4 Analysis of Naval Air Warfare Center Training Systems Division Data on the UNREP Training Task

The VETT lab at the Naval Air Warfare Center Training Systems Division (NAWCTSD) has been studying the use of VE to train Navy personnel to perform certain ship-handling maneuvers associated with the underway replenishment (UNREP) task.[21]

Among the data collected in this study are (1) the position-time tracks of the ship being controlled by the trainee in the VE and (2) the mean rating of each track obtained from a set of ratings provided by a group of experts in the UNREP task. Some sample-tracks (taken from the group of 36 tracks provided to us, together with the mean expert ratings) are shown in Figure 1. The objective of our analysis of these data was to construct a computational model that would be able to predict the mean ratings on the basis of the tracks (i.e., to construct a synthetic expert).

It was found that the ratings depended primarily on the portion of the track corresponding to the period when the ship being maneuvered by the trainee was close to the ship providing the replenishment, and only to a minor extent on the characteristics of the track during other periods. Thus, for example, the

21 M. Martin, S.W. Davidson, and S.A. Jones, "Virtual Environments in Training System Design: A Framework and Application," NAWCTSD 1998; J. Patrey, E. Sheldon, R. Breaux, and A. Mead, "Quantifying Performance of a Dynamic Shiphandling Perceptual-Action Task from Spatial Moments," NAWCTSD 1998.

efficiency of the approach phase was found to be relatively unimportant. (Overall, this result is not surprising because the cost of a collision with the replenishment ship is obviously much greater than the costs of other types of errors.) This general property of the rating procedure is well illustrated in Figure 1, where the mean rating is seen to be clearly correlated with the portion of the track in the vicinity of the replenishment ship.

A more detailed description of how the rating prediction was derived from the track information is provided in the following paragraphs.

As mentioned previously, the scoring is based on the observation that most of the interesting action during a trial is in the relatively short period of time immediately around the point of docking. Based on studying graphs as in Figure 1 above, two guidelines for a trial performance were determined. First, a good trial consists of a trial in which the subject proceeds expeditiously to (0,0), sits for two minutes, and then departs hastily. Secondly, a bad trial can occur for several reasons, but a subject colliding with or coming very near the fuel boat is always cause for a low score. Subjects who dawdle needlessly, outpace or are outpaced by the fuel boat, or drift away from the boat should also be penalized, although less drastically.

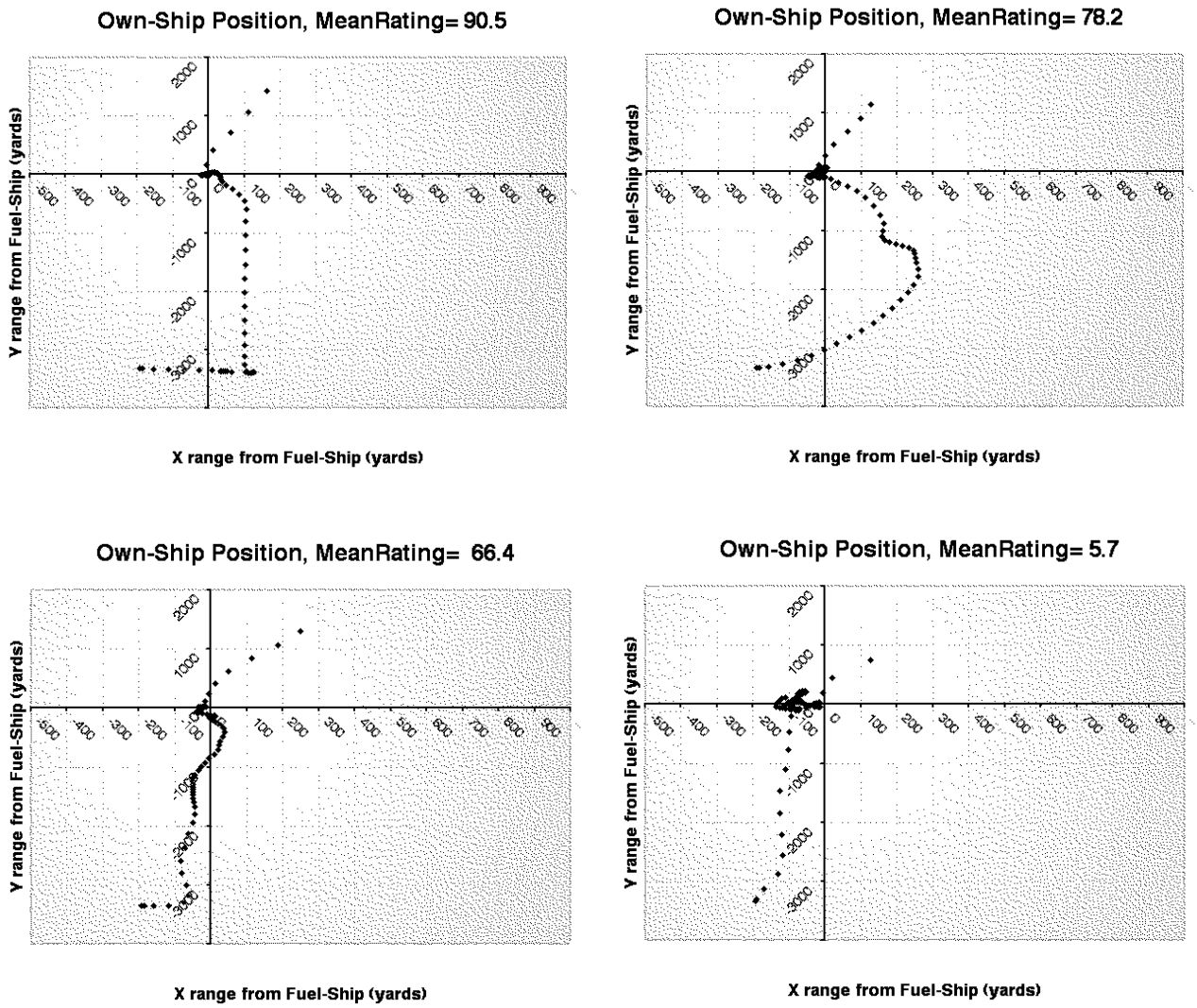

Figure 1. Four position-time tracks, together with the associated mean expert rating. Position samples taken every 15 seconds. The docking point for the replenishment fuel ship is located at the origin (coordinates x = 0 yards and y = 0 yards). Note how rating appears to depend strongly on track characteristics near replenishment ship, but only weakly on approach characteristics.

To encapsulate these observations, the following algorithm is used:

1. Extract the points from the data file during which the subject was matching speed with the fuel ship, refueling, and preparing to depart, which roughly comprised the central two minutes of each trial.
2. For each of these n points, compute a penalty function, which consists of:

$$\left.\begin{matrix} c_1 e^{c_2 x} \\ c_3 \log(x) \end{matrix}\right| \begin{matrix} x < 0 \\ x \geq 0 \end{matrix} + c_4 \log(|y|)$$

Thus, negative x positions, when the subject tended towards the fuel boat, are penalized much more severely. Constants c_1 and c_2 are chosen such that this exponential becomes very large as the subject approaches to within 20 yards of the boat.

1. Compute the base penalty P_0, which is the average of the penalty function over all n points.
2. Add to the base penalty an additional penalty proportional to n. This penalty is scaled by the scaling constant c_5, which is typically very small but large enough to penalize extreme dawdlers.
3. For each subject, compute a value X_{dev} representing the amount of x-deviation on the approach to (0,0). This value is 90-percent of the maximum x-separation between any two data points during the approach phase.
4. If X_{dev} falls above some constant value c_6, add to the subject's penalty a value proportional with constant c_7 to this x-deviation. This allows us to capture a notion of "normal approach performance" versus "bad approach" at a rough level.
5. Take the inverse of this penalty function and multiply it by constant c_8 to get the final subject score.

Thus, the final score is given by the formula:

$$c_8 \frac{1}{P_0 + c_5 n + \left.\begin{matrix} c_7 X_{dev} \\ 0 \end{matrix}\right| \begin{matrix} X_{dev} > c_6 \\ X_{dev} \leq c_6 \end{matrix}}$$

Constants c_1 through c_8 were determined programmatically through the optimizing MatLab function fmins, which uses a hill-climbing based optimization algorithm to try to match the synthetic expert score generated by the above formula to the real-world judges' scores as closely as possible.

Using this model, the correlation ρ between the mean human expert rating and the rating provided by the synthetic expert was found to be ρ = 0.94. The correlogram for these data is shown in Figure 2.

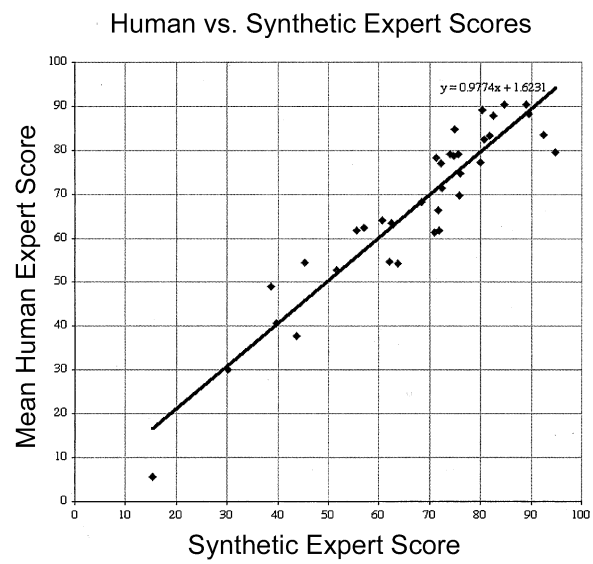

Figure 2. Correlogram showing relationship between mean human expert rating and synthetic expert rating. Best fit straight line is given by y = 0.98x + 1.62; correlation coefficient ρ is given by ρ = 0.94.

If the model is simplified by throwing out all considerations of the tracks outside the crucial two-minute period near the replenishment ship, the correlation drops to ρ = 0.87.

1.6.5 The Role of Haptics in Learning to Perform Cognitive Tasks

This work is a continuation of work described previously in Aviles[22] and in *RLE Progress Report No. 140*.[23]

As in the experiments performed previously, the basic task in the experiments performed during the past year was to memorize and recall sequences of symbols. The new experiments differed from the old

22 W.A. Aviles, *The Role of Multimodal Sensorimotor Involvement in the Learning of Cognitive Skills.* S.M. thesis, Department of Electrical Engineering and Computer Science, MIT, 1996.

in certain details that were altered in the attempt to eliminate possible artifacts. A description of the new experiments and the results obtained, together with some general comments on the whole project, are prescribed in the following subsections.

Experiment

The items to be memorized and recalled (referred to as "strings") consisted of linearly ordered sequences of letter-number pairs (referred to as "elements") drawn from the sets (A, B, C, D) and (1, 2, 3, 4). Each string contained six elements, and each set of strings, of which there were three, contained eight strings. At the beginning of each string, a stimulus label was attached. In all cases, the label consisted of the letter "S" followed by another letter. One set of strings, together with the string labels, is shown in Table 2. The task of the subject was to memorize all eight strings (together with their labels) in a set of strings and then at a later time recall the string when asked to recall it by specifying the label.

Table 2: A typical set of strings. The stimulus label is shown in bold type at the beginning of each string.

SS	A3	D2	B2	C1	B1	C3
SM	D3	A4	D1	C2	B4	B1
SP	A2	C1	D1	A3	C2	B2
SR	C4	D2	A3	A1	B1	D4
SK	D1	A2	C2	B1	D2	B2
SO	C2	B1	A4	A2	B3	B4
SF	B2	D3	C1	A4	A2	B1
SB	A3	A1	B1	C4	B2	D1

Twelve subjects, drawn from the MIT student population and ranging in age from 19-30 years, were tested under two conditions to be described below: condition V (visual) and condition VH (visual plus haptic). All subjects, who were paid for their time, participated in three experimental sessions, each of which involved only one of the two conditions and lasted roughly one hour. For six of the subjects (Group 1), session one was devoted to condition V, session two to condition VH, and session three to condition V. For the other six, the order was reversed: session one for condition VH, session two for condition V, and session three for condition VH. No subject was involved in more than one session on a given day, and the time between sessions was of the order of one to two days. The experimental design as outlined is summarized in Table 3.

Table 3: Experimental design

	Session 1	Session 2	Session 3
Group 1	Condition V	Condition VH	Condition V
Group 2	Condition VH	Condition V	Condition VH

Each of the experimental sessions consisted of a *training session* during which the subject attempted to memorize all eight strings in a given set of strings, and a *testing session* during which the subject's ability to recall the strings memorized during the training session was measured. Three distinct sets of strings were used (Set 1, Set 2, and Set 3), one for each of the three sessions in which the subject participated.

The time interval between the end of the training component and the beginning of the testing component in a session was always five minutes. In our previous work on the memorization and recall of such

23 M.A. Srinivasan et al., in *RLE Progress Report No. 140* (Cambridge: MIT Research Laboratory of Electronics, 1998), pp. 385-86.

strings, we found that the effect of varying the duration of this interval was surprisingly small, at least within the range of five minutes to 24 hours.[24]

In each training session, the subject was exposed to and examined on a given string twice before moving on to the next string in the set. After the subject had worked with each of the eight strings in this fashion, he/she was then exposed to and examined on each string over again. Thus, the subject worked with each string a total of three times in the training session.

Each time, the string was presented to the subject for a period of 15 seconds. This presentation was then followed by an exam period of 15 seconds and then a wait period of 15 seconds. Thus, each training cycle consumed 45 seconds. Because each string was involved in 3 such cycles, and there were eight strings in a set, the total duration of a training session was 18 minutes. A summary of the training session is presented in Table 4.

In order to facilitate memorization and recall, whenever a subject was exposed to a string, he/she was also shown the 4×4 matrix that displayed all possible letter number pairs (rows A, B, C, D; columns 1, 2, 3, 4). Accordingly, a typical stimulus presentation was of the form illustrated in Figure 3a.

Table 4: Outline of Training Session.

Sequence of strings:		
1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8 1,2,3,4,5,6,7,8		
For each string exposure:		
Exam: 15 Secs	Feedback: 15 secs	Wait: 15 secs
Total training time for one training session: 8 x 3 x 45 sec = 18 min		
Time between end of training session and beginning of testing session: 5 minutes		

All subjects were instructed to make use of the matrix by locating the elements of the string in the matrix and tracing out the geometric pattern corresponding to the string (see Figure 3b). In preliminary experiments, we found that presenting and using the matrix in this manner greatly improved the recall performance of most subjects. In the V condition, subjects did this tracing out solely with their eyes; their hands remained at rest on the arms of the chair in which they were seated. In the VH condition, the tracing out was performed manually (with the index finger) as well as visually.

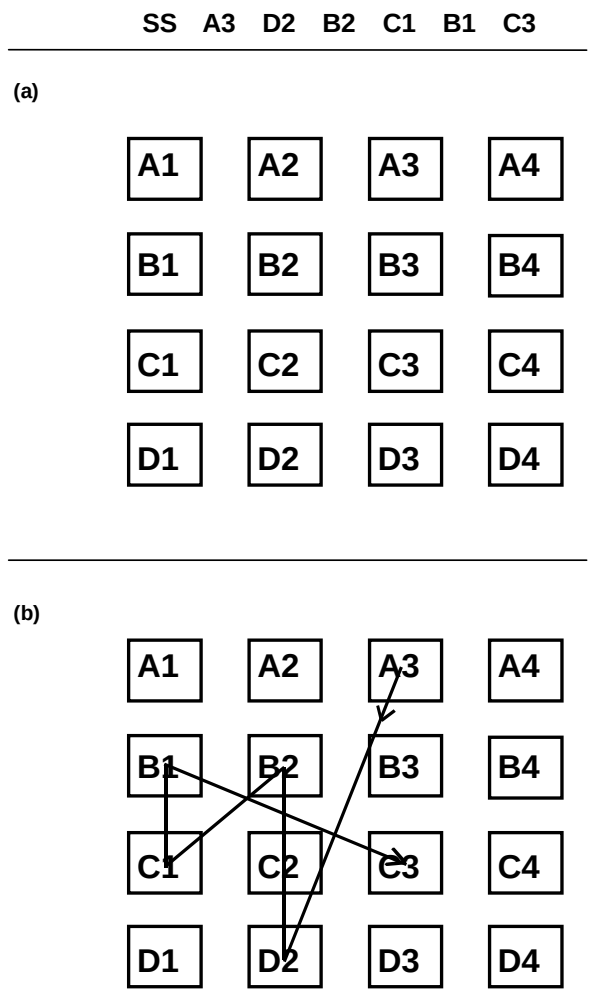

Figure 3. Inclusion of matrix with string. (a) presentation; (b) tracing out string on matrix.

In the V condition, the subjects were required during the exam component of the training session to write down the string to which they had been exposed (including the string label) and then grade themselves by accessing the correct version of the string and using it to score the string estimate they had written down. In order for an estimated string element to be scored as correct, both characters in the element

24 W.A. Aviles, *The Role of Multimodal Sensorimotor Involvement in the Learning of Cognitive Skills.* S.M. thesis, Department of Electrical Engineering and Computer Science, MIT, 1996.

had to be correct and the element had to occur in the proper position of the string. The training in the VH condition was similar, except for the above-mentioned manual tracing out of the string on the matrix. The subject not only performed this manual tracing during the 15 second exposure period, but also recorded the estimated string during the exam period by tracing out the string on the matrix with a pencil.

After a subject completed the training session, recall was measured in the testing session using a forced-choice written exam. In this exam, the subject was provided with a sequence of eight string labels and asked to provide the strings corresponding to the labels. In the V condition, the subject was required to specify the string by writing the string down as a linear sequence of elements (as in the training session for the V condition). In the VH condition, the subject was required to specify the string by manually tracing it out on the matrix (as in the training session for the VH condition).

The presentation order of string labels in the written test was random and each of the eight string labels appeared exactly once. The subjects were instructed to recall the appropriate strings to the best of their abilities and guess when they were unsure. No feedback was presented during these tests and the scoring method used was the same as in the training sessions.

In addition to these objective tests, subjects were asked to fill out a questionnaire at the end of the last experimental session in an attempt to gain insight into individual memorization/recall strategies and the effects of condition VH versus condition V.

Results

As indicated previously, the results of the written examinations performed during the testing sessions were scored in the same fashion as the results in the training sessions. In order to be marked correct, a string element had to be completely recalled (i.e., both characters had to be correct) and had to be in the proper position within the string. A single point was awarded for every string element correctly recalled. Incorrect answers received zero points. The forced-choice paradigm ensured that there were no blanks left on the subject's examination sheet.

The results for both groups, presented as percent correct, are shown in Figure 4. Individual subject scores for each stimulus set are shown, as well as the average scores across subjects. Standard error bars are included with the average scores.

The individual-subject data show that the intersubject variability is large for both groups. While some subjects achieve perfect or near perfect scores on some tests (S2, S3, S6, S8, S11, S12 on the tests in the third session), others never obtain scores above 30% (S5, S7). The average scores for both groups show that there is a strong learning effect across sessions. For subjects in Group 1 (upper panel), the average score across subjects for the first session (condition V) was 31%. In the second session, where condition VH was used, the average score increased to 47%. However, in the third session, where condition V was tested again (with a new set of strings), the average score increased further to 65%. Data for subjects in Group 2 (lower panel) show a similar pattern. For this group, where the conditions were presented in reverse order (first VH, then V, then VH), average scores across subjects for the three sessions were 26%, 39%, and 67% respectively.

Averaged over both groups and sessions (as well as subjects), the overall average scores were 45% and 47% for V and VH, respectively.

In summary, the results of this experiment indicate that (1) using haptic involvement as an aid in this memory task did not provide a benefit over performance with vision alone and (2) a learning effect was seen in both conditions.

The data collected from questionnaires and personal interviews with the subjects showed that most subjects felt that the use of haptic involvement aided their recall performance, even though our test results did not demonstrate this. They also stated that they had greater difficulty attaching the string to the stimulus label than they did remembering the string sequence. Again, this did not show up in our test results; in no case did the subject record the string elements correctly but attach the string to an incorrect label.

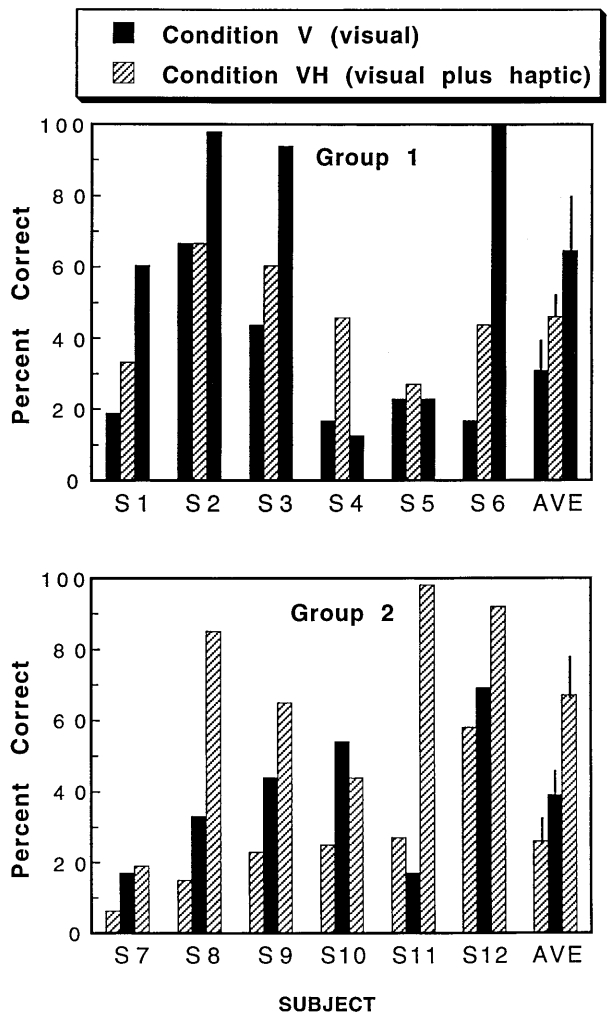

Figure 4. Results of testing sessions.

Discussion

The negative results we obtained in this version of the experiment are essentially the same as those we obtained in previous versions. These previous versions include not only the version described in Aviles,[25] but a variety of additional versions explored in the interim. Some of the versions were changed in an attempt to reduce intersubject variability. For example, some earlier versions did not use a forced-choice paradigm; the subject was allowed the option of not responding to a particular test item. We went to the forced-choice paradigm in order to eliminate variability associated with the decision of whether or not to respond. Also, some versions intermixed the V and VH conditions during the same session rather than evaluating the conditions separately. Moreover, some variations of the experiment did not require that the subjects be tested in the same mode in which they were trained. In other words, independent of the training condition (V or VH), the test session always required the subject to specify the string by writing it down as a linear sequence of elements (as in the V condition). Other changes were made to eliminate factors that were thought to be deleterious. For example, earlier test sessions were three hours in duration with short breaks. Subjects commented during interviews that they became drowsy during the sessions. Our later sessions were all one hour in duration to eliminate the possibility that fatigue played a role in performance. As indicated previously, none of these changes altered the basic results: in no case, did haptic involvement change recall performance.

Although we certainly cannot prove it, based on the results that we have obtained in all the versions that we have considered, we believe that the results reported in this do not represent merely artifacts of experimental procedure. Despite this belief, however, we are not comfortable with the results. We are under the impression that haptic involvement *does* aid in the recall of certain analogous “strings” used in everyday life. For example at least intuitively, it does seem to aid in the recall needed to successfully “punch in” telephone numbers, combination lock sequences, security-system passwords, etc.

Among the possible explanations for the apparent contradiction between this impression and our experimental results are the following.

First, perhaps our impression is false. We have not yet found any formal experimental data to support this impression.

Second, even if our impression is correct, maybe the benefit of the haptic involvement in the real world cases would disappear if one (1) confined one’s attention solely to 2-D displays perpendicular to gaze direction (certainly, the visual system is good at extracting and storing information from such displays) and (2) gave the subjects explicit instructions about how to best make use of such visual displays (analogous to our instructions to visually trace out the string on the matrix).

25 W.A. Aviles, *The Role of Multimodal Sensorimotor Involvement in the Learning of Cognitive Skills*, S.M. thesis, Department of Electrical Engineering and Computer Science, MIT, 1996.

Third, maybe somehow the sensations associated specifically with the "punching in" operation are important. Haptically, the punching in operation differs from the tracing operation in that the former operation (1) includes movement in depth as well as in the lateral dimensions and (2) involves the production and feedback of force.

Fourth, and finally, it is possible that the time constants (related to consolidation and decay) associated with visual memory are very different from those associated with haptic memory, and that the benefit of adding haptic involvement to the visual involvement varies strongly with temporal factors that have not yet been explored in our studies.

In any case, it is quite possible that the benefit of haptic involvement would become apparent if a different task or training procedure were chosen. For example, if the tracing-out memory aid were exploited in the haptic case but not the visual case, the relative scores of VH over V would improve. However, we regard this as "cheating"; obviously, the relative scores could be increased by degrading performance in the V condition. A similar, but somewhat more legitimate path, would be to compare conditions V and VH under conditions in which there is visual interference or a large visual work load. Such conditions often occur in important practical situations. A more interesting path, however, would be to explore situations in which the haptic involvement involved not only position and changes in position (as in the tracing-out operation) but also exertion and perception of force, as well perhaps as the sensing of texture. At present, we know very little about the extent to which haptic involvement of these kinds would aid in learning to perform cognitive tasks.

1.6.6 Conveying the Touch and Feel of Virtual Objects

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 maturing and is opening 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. For recent reviews of haptic interfaces, see Srinivasan (1995) and Srinivasan and Basdogan (1997).[26]

Two major devices for performing psychophysical experiments, the linear and planar 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 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.

Haptic Rendering Techniques: Point and Ray-Based Interactions

Haptic rendering, a relatively new area of research, is concerned with real-time display of the touch and feel of virtual objects to a human operator through a force reflecting device. It can be considered as a subdiscipline of computer haptics. A major component of the rendering methods developed in our laboratory is a set of rule-based algorithms for detecting collisions between the generic probe (end-effector) of a force-reflecting robotic device and objects in VEs. We use

26 M.A. Srinivasan. "Haptic Interfaces, In Virtual Reality: Scientific and Technical Challenges," In *Report of the Committee on Virtual Reality Research and Development*, N.I. Durlach and A.S. Mavor, eds. (Washington, D.C.: National Research Council, National Academy Press, 1995); M.A. Srinivasan and C. Basdogan, "Haptics in Virtual Environments: Taxonomy, Research Status, and Challenges," *Comput. Graph.* 21(4): (1997), which won the 1997 *Computers and Graphics* Best Paper Award.

a hierarchical database, multithreading techniques, and efficient search procedures to reduce the computational time and make the computations almost independent of the number of polygons of the polyhedron representing the object. Our haptic texturing techniques enable us to map surface properties onto the surface of polyhedral objects.

Two types of haptic rendering techniques have been developed: point-based and ray-based. In point-based haptic interactions, only the end point of the haptic device, also known as the end-effector point or haptic-interface point (HIP), interacts with objects. Since the virtual surfaces have finite stiffnesses, the end point of the haptic device penetrates into the object after collision. Each time the user moves the generic probe of the haptic device, the collision detection algorithms check to see if the endpoint is inside the virtual object.[27] In ray-based haptic interactions, the generic probe of the haptic device is modeled as a finite ray whose orientation is taken into account, and the collisions are checked between the ray and the objects.[28] Both techniques have advantages and disadvantages. For example, it is computationally less expensive to render 3-D objects using the point-based technique. Hence, we achieve higher haptic servo rates. On the other hand, the ray-based haptic interaction technique handles side collisions and can provide additional haptic cues for conveying to the user the shape of objects.

Journal Article

Ho, C.H., C. Basdogan, and M.A. Srinivasan. "Efficient Point-based Rendering Techniques for Haptic Display of Virtual Objects." *Presence: Teleop. Virt. Environ.* Forthcoming.

1.6.7 Constructing Multimodal Virtual Environments

In order to develop effective software architectures for multimodal VEs, we have experimented with multi-threading (on Windows NT platform) and multi-processing (on UNIX platform) techniques and have successfully separated the visual and haptic servo loops. Our experience is that both techniques enable the system to update graphics processing at almost constant rates, while running the haptic process in the background. We are able to achieve good visual rendering rates (30 to 60 Hz), high haptic rendering rates (more than 1 kHz), and stable haptic interactions. Although creating a separate process for each modality requires more programming effort, it enables the user to display the graphics and/or haptics on any desired machine(s), even those in different locations, as long as the physical communication between them is provided through a cable. Programming with threads takes less effort, but they are not as flexible as processes.

We have also developed a graphical interface that enables a user to construct virtual environments by means of a user-defined text file, toggle stereo visualization, save the virtual environment, and quit from the application. This application program was written in C/C++ and utilizes the libraries of (1) Open Inventor (from Silicon Graphics, Inc.) for graphical display of virtual objects, (2) ViewKit (from Silicon Graphics, Inc.) for constructing the graphical user interface (e.g., menu items, dialog boxes, etc.), and (3) parallel virtual machine (PVM), a well-known public domain package, for establishing the digital communication between the haptic and visual processes. The user can load objects into the scene and assign simple visual and haptic properties to the objects using this text file. Following the construction of the scene using the text file, the user can interactively translate, rotate, and scale objects, and the interface will automatically update both the visual and haptic models.

Using the haptic rendering techniques and the user interface described above, we have designed experiments to investigate human performance involving multimodal interactions in virtual environments. The user interface enabled several experimenters to rapidly load virtual objects into desired experimental scenarios, interactively manipulate (translate, rotate, scale) them, and attach sophisticated material and visual properties to the virtual objects.

Generating Multimodal Stimuli

Once the software and hardware components were put together for integrating multiple modalities, we focussed on developing techniques for generating multimodal stimuli. Our interest in generating multimodal stimuli is two-fold: (1) we would like to develop

27 C.H. Ho, C. Basdogan, and M.A. Srinivasan, "Haptic Rendering: Point- and Ray-based Interactions," *Proceedings of the Second PHANToM User's Group Workshop,* Cambridge, Massachusetts, October 1997.

28 C. Basdogan, C.H. Ho, and M.A. Srinivasan, "A Ray-based Haptic Rendering Technique for Displaying Shape and Texture of 3D Objects in Virtual Environments," DSC-Vol. 61, *Proceedings of the ASME Dynamic Systems and Control Division*, G. Rizzoni, ed. (ASME, 1997), pp. 77-84.

new haptic rendering techniques to display shape, texture, and compliance characteristics of virtual objects, and (2) utilize these techniques in our experiments on human perception and performance to study multimodal interactions. Our progress in this area is summarized under two headings: texture and compliance.

Texture

Since a wide variety of physical and chemical properties give rise to real-world textures, a variety of techniques are needed to simulate them visually and haptically in VEs. Haptic texturing is a method of simulating surface properties of objects in virtual environments in order to provide the user with the feel of macro and micro surface textures. Previously, we had developed two basic approaches: force perturbation, where the direction of the displayed force vector is perturbed, and displacement mapping, where the microgeometry of the surface is perturbed.

Using these methods, we have successfully displayed textures based on Fourier series, filtered white noise, and fractals. But the display of haptic textures using the force perturbation technique was effective only in a certain range (0.5 mm to 5.0 mm in height). To extend the range of haptic textures that can be displayed, we have modified the algorithm to include the calculation of the location of the point closest to the object surface prior to collision detection. Using this additional information, we are now able to render macro-textures (> 5.0 mm height) as well. We have also experimented with 2-D reaction-diffusion texture models used in computer graphics and successfully implemented them for haptics to generate new types of haptic textures. The reaction-diffusion model consists of a set of differential equations that can be integrated in time to generate texture fields. Moreover, we have developed techniques to extend our work on 2-D reaction-diffusion textures to three-dimensional space. We have also studied some of the image and signal processing techniques frequently used in computer graphics to convolve 2-D images of spots (i.e., simple 2-D geometric primitives such as circles, squares, and triangles) with noise functions in order to generate a new class of haptic textures.

In summary, the following texture rendering techniques have been developed: (1) force perturbation, (2) displacement mapping. Using these rendering techniques, we can display the following types of synthetic haptic textures: (1) periodic and aperiodic haptic textures based on Fourier series approach, (2) noise textures (based on the filtered white noise function), (3) fractal textures, (4) reaction-diffusion textures (a set of differential equations are solved in advance to generate a texture field that can be mapped onto the 3-D surface of the object), and (5) spot-noise textures (the noise function is convolved with 2-D images of spots to generate distorted spots that can be displayed haptically). In addition, we have developed image-based haptic textures (the grayscale values of an image are used to generate texture fields that can be mapped onto the surface of 3-D objects).

The techniques described above enable the user to create and display synthetic textures. In addition, we utilized our haptic device, the PHANToM, as a haptic recording tool to acquire information about the nature of real textures. Then this information is combined with texture rendering techniques to playback textured surfaces in virtual environments. To sample real-life textures, the PHANToM makes a single stroke on a textured surface by following a given trajectory and exerting a constant normal force. The force, velocity and positional data are recorded and analyzed. We have developed a stick-slip friction model in which the coefficients of friction are derived from the force and velocity data digitized from the actual texture. These friction coefficient values are then used as parameters governing the playback of virtual textures.

Compliance

We have developed procedures for simulating compliant objects in virtual environments. The developed algorithms deal directly with geometry of 3-D surfaces and their compliance characteristics, as well as the display of appropriate reaction forces, to convey to the user a feeling of touch and force sensations for soft objects. The compliant rendering technique has two components: (1) the deformation model to display the surface deformation profile graphically; and (2) the force model to display the interaction forces via the haptic interface. The deformation model estimates the direction and the amount of deformation (displacement vector) of each node (i.e., a vertex) of the surface when it is manipulated with the generic probe of the haptic interface device. We utilize a polynomial model or a spline-based model to compute the displacement vector of each node and to visually display deformations. In the force model, a network of springs is utilized to compute the direction and magnitude of the force vector at the node that is closest to the contact point. The techniques

described here enable the user to interactively deform compliant surfaces in real-time and feel the reaction forces.

1.6.8 Experimental Studies on Interactions Involving Force Feedback

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 effective haptic interfaces. In addition, 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 provides not only important background for VE design, but the availability 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 previously 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.[29] However, the influence of contact sounds on the perception of object stiffness is not as dramatic when tapping virtual objects through a haptic interface.[30] Over the past year, we have designed and conducted more experiments to explore the human haptic resolution as well as the effect of haptic-auditory and haptic-visual interactions on human perception and performance in virtual environments.

Haptic Psychophysics

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 over the past few years with a haptic interface to measure human haptic resolution in discriminating fundamental physical properties of objects through active touch. A computer controlled electromechanical apparatus, called the linear grasper, 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% for stiffness, 12% for viscosity, and 20% 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. This hypothesis, called 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. A special case of this hypothesis is that when humans discriminate stiffness, viscosity or mass, they do so by discriminating the mechanical work needed for actually deforming the objects. Implications of these results to 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 discriminable objects.

In preparation for our planned experiments to explore the interaction and separability of haptically-displayed stimulus qualities (stiffness, viscosity, mass) and further test the work hypothesis developed to

29 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 (ASME, 1996), pp. 555-59.

30 D.E. DiFranco, G.L. Beauregard, and M.A. Srinivasan, “The Effect of Auditory Cues on the Haptic Perception of Stiffness in Virtual Environments,” *Proceedings of the ASME Dynamic Systems and Control Division*, DSC-Vol. 61, G. Rizzoni, ed. (ASME, 1997), pp. 17-22.

explain results of previous experiments, we have refurbished the hardware of the linear grasper. This work included repair and rewiring of the position, force, and acceleration sensor systems; construction of a new regulated power source; and incorporation of additional (analog) low-pass filtering of the signal from the acceleration sensor. Software has been written for calibration of the sensors and for testing the system. This software is able to control the hardware to enable simulation of arbitrary combinations of mass, viscosity, and stiffness.

Haptic-Auditory Interactions

In this series of experiments, we investigated the effect of the timing of a contact sound on the perception of stiffness of a virtual surface. The PHANToM, a six degree of freedom haptic interface with 3 degrees of active force feedback, was used to display virtual haptic surfaces with constant stiffness. Subjects heard a contact sound lasting 130 ms through headphones every time they touched a surface. Based on our earlier work on stiffness discrimination, we initially hypothesized that presenting a contact sound prior to actual impact creates the perception of a less stiff surface, whereas presenting a contact sound after actual impact creates the perception of a stiffer surface. However, the findings indicate that both pre-contact and post-contact sounds result in the perceptual illusion that the surface is less stiff than when the sound is presented at contact.

Haptic-Visual Interactions

Previously we have shown how the perception of haptic stiffness is influenced by the visual display of object deformation.[31] 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 perceived by the subject can be effectively increased by reducing the visual deformation of the object.

In continuing this line of investigation on how vision affects haptic perception, we designed two sets of experiments to test the effect of perspective on the perception of geometric and material properties of 3-D objects. Virtual slots of varying length and buttons of varying stiffness were displayed to the subjects, who then were asked to discriminate their size and stiffness respectively using visual and/or haptic cues. The results of the size experiments show that under vision alone, objects that are farther away are perceived to be smaller due to perspective cues and the addition of haptic feedback reduces this visual bias. Similarly, the results of the stiffness experiments show that compliant objects that are farther away are perceived to be softer when there is only haptic feedback and the addition of visual feedback reduces this haptic bias. Hence, we conclude that our visual and haptic systems compensate for each other such that the sensory information that comes from visual and haptic channels is fused in an optimal manner. In particular, the result that the farther objects are perceived to be softer when only haptic cues are present is interesting and perhaps suggests a new concept of *haptic perspective*. To ensure that this result was not an artifact of the robot arm (i.e., position and force errors due to the kinematics of the haptic device) or our experimental design, we performed three different tests, but the results remained the same.

Meeting Paper

Wu, W., C. Basdogan, and M.A. Srinivasan. “The Effect of Perspective on Visual-Haptic Perception of Object Size and Compliance In Virtual Environments.” To be presented at the *Winter Annual Meeting of ASME*, Nashville, Tennessee, November 14-19, 1999.

1.6.9 Preliminary Investigations on Expanding the Perceived Workspace

A larger haptic workspace would be useful in several haptic applications. Unfortunately, the physical haptic workspace that is provided by the currently available haptic devices is limited due to several reasons (e.g., the dimensions of the mechanical links of the device for effective force feedback and the distance reachable by the human user). We are currently working on the concept of expanding the perceived haptic workspace using our rendering algorithms and the user interface. We briefly summarize our approach here on how the perceived space can be extended.

31 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, ASME, 1996, pp. 555-59.

Currently, we are testing the feasibility of these approaches for integration into our existing multimodal software.

Scaling the Cursor Movement

If the cursor movements, determined by the end point of the haptic device, are scaled appropriately before they are displayed in the visual space, then the haptic space may be perceived as bigger than the actual one. We define a scaling ratio as the size of visual workspace to the size of haptic workspace (SVW/SHW), and we set its value higher than one to make sure that the haptic workspace perceived by the human user becomes bigger than its physical dimensions. Hence, the user will feel as if he/she is traveling faster in the visual workspace as a result of this scaling.

Scaling the Visual Size of Stylus

Similarly, the visual size of the stylus can be scaled by the ratio of workspaces to extend the reachable visual workspace. In the past, we have developed a ray-based rendering technique to haptically interact with 3-D objects in virtual environments. In this technique, the stylus of the haptic interface device is modeled as a line segment and the collisions are detected between the line segment and the 3-D objects in the scene. Since our ray-based rendering technique relies on detecting collisions between a line segment model of the stylus and the 3-D objects and it can handle side collisions, small movements of the hand in haptic space will result in the display of longer stylus movements in visual workspace. This can be imagined by assuming that a blind person is exploring the surface of a 3-D object using a special stick that can be extended automatically as he/she presses a button on the stick to feel the objects that are further away.

1.6.10 Haptics Across the World Wide Web

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 previously using the PHANToM haptic interface was developed to be used across the World-Wide-Web. The program was written in Java, using multi-threading 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. Users are asked to (1) discriminate the stiffness of sets of two springs, displayed visually on the screen and haptically with the PHANToM, and (2) 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 multimodal VEs across the internet.

1.6.11 The Role of Haptics in Shared Virtual Environments

We are conducting a set of human experiments to investigate the role of haptics in shared virtual environments (SVEs). Our efforts are aimed at exploring (1) whether haptic communication through force feedback can facilitate a sense of togetherness between two people at different locations while interacting with each other in SVEs. If so, (2) what types of haptic communication/negotiation strategies they follow, and (3) if gender, personality, or emotional experiences of users can affect the haptic communication in SVEs. The experiment concern a scenario where two people, at remote sites, cooperate to perform a joint task in a SVE. The experiments are abstractions from real situations to create a more controlled environment suitable for explanatory studies in the laboratory. During the experiment, subjects were not allowed to meet their remote partner, and did not know where their partner was located. The participants were in different rooms but saw the same visual scene on their monitor and felt the objects in the scene via a force feedback device, the PHANToM.

The goal of the task was to move a ring with the help of another person without touching a wire (see Figure 5). A ring, a wire, and two cursors attached to the ring were displayed to the subjects. Haptic interactions between cursors as well as between cursor and the ring were modeled using a spring-damper system and a point-based haptic rendering technique.[32] Subjects were asked to move the ring back and forth on the wire many times, in collaboration with each other such that contact between the wire and the ring was minimized or avoided. If the ring touched the wire, the colors of the ring and the surrounding walls were changed to red to warn the subject of an error. They were changed back to their original colors when the subjects corrected the position of the ring. To hold the ring, both subjects needed to press on the ring towards each other above a threshold force. If they did not press on the ring at the same time, the ring did not move and its color was changed to gray to warn them. To move the ring along the wire, they each needed to apply an additional lateral force.

Two sensory conditions have been explored to investigate the effect of haptic communication on the sense of togetherness: (1) both visual and haptic feedback provided to the participants, and (2) only visual feedback was provided to the participants

Performance and subjective measures were developed to quantify the role of haptic feedback in SVEs. Performance measure was derived from the following measurements: (1) total amount time takes to complete task, (2) the ratio of erroneous-time to total time. Several *subjective* questions were asked, through a questionnaire, in four categories including their (1) performance, (2) their sense of "being together," (3) emotional reactions, and (4) personality profile. Each of the questions in categories 1, 2, and 3 were rated on a 1-7 scale. Subjective measures were correlated with the performance measures to deduce conclusions on the effect of haptic feedback to the task performance and the sense of being with someone in SVEs. The results suggest that haptic feedback significantly improves the performance and contributes to the feeling of “sense of togetherness” in SVEs.[33]

Journal Articles

Durlach, N.I., and M. Slater. “Presence in Shared Virtual Environments and Virtual Togetherness.” *Prescence: Teleop. Virt. Environ. Forthcoming*.

Meeting Papers

Basdogan C., C. Ho, M. Slater, and M.A. Srinivasan. “The Role of Haptic Communication in Shared Virtual Environments.” Paper published in the *Proceedings of the Third PHANToM Users Group Workshop*, Dedham, Massachusetts, October 3-4, 1998.

Ho, C., C. Basdogan, M. Slater, N.I. Durlach, and M.A. Srinivasan. “An Experiment on The Influence of Haptic Communication on the Sense of Being Together.” Paper presented at the British Telecom Workshop on Presence in Shared Virtual Environments, Ipswich, UK, June 10-11, 1998. (http://www.cs.ucl.ac.uk/staff/m.slater/BTWorkshop/touchexp.html).

Additional Publications

In addition to the publications cited in connection with the work described above, the following have also been published or are forthcoming.

Durlach, N.I. “Virtual Reality.” In *Encyclopedia of Psychology*. New York: Oxford University Press. Forthcoming.

Reed, C.M., and N.I. Durlach. “Note on Information Transfer Rates in Human Communication.” *Prescence: Teleop. Virt. Environ.* 7: 509-18 (1998).

32 C.-H. Ho, C. Basdogan, M. Slater, N.I. Durlach, and M.A. Srinivasan, “An Experiment on The Influence of Haptic Communication on the Sense of Being Together,” paper presented at the British Telecom Workshop on Presence in Shared Virtual Environments, Ipswich, Massachusetts, June 10-11, 1998. (http://www.cs.ucl.ac.uk/staff/m.slater/BTWorkshop/touchexp.html);C.-H. Ho, C. Basdogan, and M.A. Srinivasan, “Haptic Rendering: Point- and Ray-based Interactions,” *Proceedings of the Second PHANToM User's Group Workshop,* Cambridge, Massachusetts, October 1997.

33 C. Basdogan, C. Ho, M. Slater, and M.A. Srinivasan, “The Role of Haptic Communication in Shared Virtual Environments,” paper published in the *Proceedings of the Third PHANToM Users Group Workshop*, Dedham, Massachusetts, October 3-4, 1998.

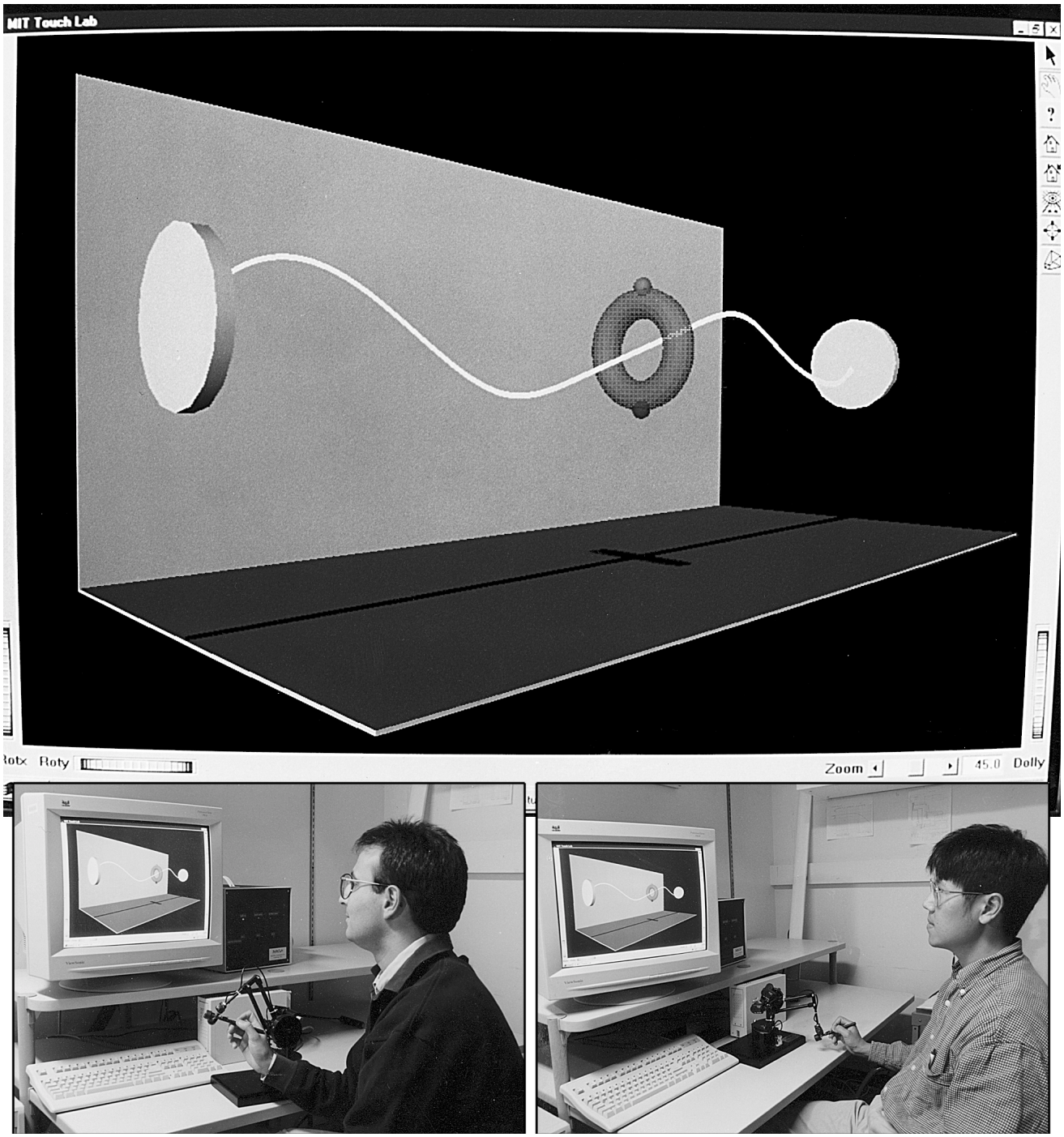

Figure 5. Sense of presence in shared virtual environments. (Bottom left: Research Scientist Dr. Cagatay Basdogan; right: graduate student Chih-Hao Ho.)

1.7 Training for Remote Sensing and Manipulation

Sponsor

U.S. Navy - Office of Naval Research
Subcontract 40167[34]

Project Staff

Nathaniel I. Durlach, Dr. Thomas E.v. Wiegand, Dr. David W. Schloerb, Rebecca L. Garnett, Matthew E. Esch, David H. Manowitz, Jonathan L. Zalesky

1.7.1 Introduction

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 pilots of teleoperated mini-submarines, often called remotely operated vehicles (ROVs). Previous work on this project is described in previous *Progress Reports*.[35]

During the past year, the MIT TRANSoM effort has included work in four areas: (1) tether information, (2) sonar simulation, (3) auditory display of thruster state, and (4) force-feedback training transfer. Work on these areas is described in the following subsections.

1.7.2 Tether Information

One of the main problems in piloting a remotely controlled, tethered, underwater vehicle (ROV) is knowledge about the state of the tether. Not only can the tether exert forces on the vehicle that modify its dynamics, but the tether can become entangled with itself or with objects in the underwater environment. The pilot has to create a mental model of the position and shape of the tether (and the forces it may exert) based on knowledge of its length, the ROV's present position and orientation, the ROV's past behavior, the direction and magnitude of the water current, and the physical properties of the tether (e.g., its stiffness and buoyancy). Along with everything else the pilot has to do, constructing this mental model constitutes a significant challenge.

Theoretically, this problem can be attacked in two ways: (1) by developing a performance aid that would acquire and display tether information to the pilot during operation of the ROV and (2) by developing a training aid that would display simulated tether behavior to the pilot while being trained to pilot an ROV in a virtual environment designed for pilot training. In each case, the value of the aid would be judged by the extent to which it improved pilot performance in the real world.

Tether Models

The VE system we use incorporates a three-dimensional space with obstacles located at various positions. The velocity vector of the current is assumed to be locally constant in space and time, although it can change during a given scenario. The effects of the environment on the ROV itself have already been modeled. What is needed is a model of how the ROV and the environment affect the tether and how the tether acts upon the ROV.

To be useful, the tether model must meet certain requirements. It must produce as output (1) the position of the tether, which may consist of a number of connected links, and (2) the magnitude and direction of the force applied by the tether to the ROV. The inputs to our models provided by the VE include the tether's endpoints, the direction and magnitude of the current, the direction and magnitude of the ROV's velocity, the length of tether deployed, and certain properties of the tether. The properties of the tether needed by the models we examined are the diameter of the tether, its weight or lift underwater, drag coefficients with the flow of water, modulus of elasticity, and the bending radius at which it is stored.

Dynamic tether models include the effects of the movement of the ROV and the tether when finding the tether's current position. This addition makes these models very complex because the velocity of the tether along its length must be found in addition to its position and tension. These models generally make very few assumptions about the tether environment and neglect the effect of bending biases. In addition, they assume that the tether is relatively uniform in diameter and that the current effects are well-

34 Imetrix Corporation, Cataumet, Massachusetts.

35 N.I. Durlach et al., "Training for Remote Sensing and Manipulation," *Progress Report No.139* (Cambridge, Massachusetts: MIT Research Laboratory of Electronics, 1997) pp. 388-95; N.I. Durlach et al., "Training for Remote Sensing and Manipulation," *Progress Report No. 140* (Cambridge, Massachusetts: MIT Research Laboratory of Electronics, 1998), pp. 393-98.

behaved (no waves or vortices). Each of the models considered below also has its own additional assumptions.

The general method of solution used in these models is to balance all forces acting at discrete points, connected by stiff links, along the cable. The variables solved for in these models, using finite difference schemes, are the tension in the tether, magnitude and direction of the velocity of the tether, and angle between the links at each discrete point. The forces acting on the tether that are considered include the weight or buoyancy of the tether, a fluid inertial force, and the drag force due to the current flow. These forces are set equal to the mass of the cable times the acceleration in the different coordinate directions. Each dynamic model may include or neglect additional interactions. Also, all the dynamic models place the tether in natural coordinates, i.e., the coordinate system is centered around each individual segment of tether. This allows for easy representation of the forces acting on the tether.

In static tether models, the velocities of the ROV and the tether are assumed to be much less than the current and are not included in determining the position of the tether. This assumption greatly simplifies the equations governing the tether and allows it to be defined solely in terms of its position and tension. In many cases, this assumption is valid because, in fact, the ROV velocity is small compared to that of the current. In general, the accuracy of this assumption depends strongly on the type of ROV mission being simulated. If 1000 meters of tether is deployed, as in some deep ocean missions, and the area in which the ROV is operating is small, then the effects of the movement of the ROV will be minimal. However, if the site of exploration is a small bay and only 50-100 meters of tether is deployed, then the velocity of the ROV may begin to have some important effects.

In addition to the assumption that the situation is relatively static, the models that we considered assume that the tether is relatively uniform in diameter and that the current effects are well-behaved (no waves, no vortices). Bending stiffness or stretching of the tether was also not included.

Smart Tether Information Acquisition

A tether performance aid would require the integration of a tether model with a tether sensing system. In this section, we consider the task of sensing tether shape and, in particular, the use of strain-gauge technology to estimate this shape. Strain is defined as the ratio of the change in length of a material to its initial unstressed length. Axial strain occurs when the two ends of the tether are pulled apart. Since we do not want this type of strain to enter into our estimates of shape, sensors have to be positioned in a manner that allows compensation for this strain. Bending strain occurs in the tether when the tether is forced to a particular angle or radius of curvature. In such a situation, one side of the tether experiences an elongation while the opposite side experiences a compression. It is this strain differential that is to be measured. Torsional strain occurs when the tether is twisted about its longitudinal axis. Measurements of this type of strain are needed since they affect how the sensors are oriented in the global coordinate frame.

Estimates of strain by means of strain gauges, which convert strain to changes in electrical resistance, can be made more sensitive by employing a Wheatstone Bridge. The total strain, which is represented by a change in Vout, equals the sum of four individual strains. If all four strains were equal, the total would be zero, and Vout would remain unchanged (and equal to one-half of the regulated DC voltage). In general, the bridge network gives the sensor system greater sensitivity and resolution than that which is provided by a single strain gauge. The desirability of using two rather than four gauges depends on the level of noise in the electrical components connecting the sensors and on the amount of sensitivity required. The amount of electrical noise in the sensor system will need to be determined empirically.

Integrated circuits are available to incorporate analog to digital conversion with on-chip serially-based communication protocols, allowing the analog outputs of a long line of sensors to be individually selected. With such a serial technique, only a few electrical wires are needed to run the length of the cable; each sensor does not need its own communication line. Although it may be possible to use other types of sensors in the same way, development of alternative systems would probably require significant research.

Though strain gauges are usually employed to measure the strain on stiff pieces of metal, their specifications meet the requirements of a thin, relatively flexible cable. The minimum bend radius required is determined by the radius of the tether cable. Since most tethers have a radius of at least 1 cm, the specification of 0.3 mm for the minimum bend radius is more than adequate. The maximum elongation or

compression for tethers with copper wire (the limiting factor) is about 0.5%, a magnitude well below the specification for strain gauges. Since most tethers use copper wire, we assume that a maximum elongation of 0.5% is normal for most tethers. Since the circumference of a tether with a radius of 1 cm is a little greater than 6 cm, it is feasible to place a few 6 mm strain gauges on the exterior of the tether.

Part-task Tether Training System

The primary purpose of the part-task trainer is to serve as a tool for performing isolated tests involving tether-awareness training. This is an important cognitive skill for ROV operators that allows them to mentally track the position and shape of the tether in relation to other objects in the environment. In addition, the part-task trainer serves as a simplified test bed for verifying new tether model designs. The part-task trainer's design also allows for extending its functionality to include other aspects of the underwater ROV training environment, such as providing auditory feedback about the ROV's thruster status or the force effects of tether drag.

The current part-task trainer provides only a two-dimensional representation of the underwater environment. In a typical simulation, the main display depicts an overhead view of the underwater environment. It appears as a flat grid representing the underwater environment with horizontal but no vertical maneuverability. The third dimension was intentionally omitted to simplify the simulation. Models developed and tested in the part-task trainer can be converted to three-dimensional models when needed.

An ROV and a variable number of obstacles can be placed in the underwater environment. Colored polygons portray the ROV and obstacles. The tether connects the ROV to its home base. In practice, the tether would be connected to either a ship or a platform. The home base is shown in the trainer by a small black square. The tether is drawn as a series of connected white line segments running between the home base and the ROV.

Several displays are provided to give the user ancillary information about the environment. These displays are the water current display, tracers, and the tether information display. The water current display resides in the upper left corner of the trainer view window. It is a white circle with an arrow and resembles a compass. The orientation of the arrow shows the direction of the water current flow, and the length of the arrow indicates the magnitude of the current's flow.

A tracer is a line segment indicating the position and heading of the ROV at a specific point in time. A new tracer is plotted at specified uniform time intervals. The tracers give the driver information about the history of the positions and headings of the ROV. The trainer is set to keep track of a default number of 100 tracers, and the default tracer recording time interval is one second.

The tether information display serves as both a debugging tool for the tether models and an information source for the driver. The tether information display can provide the length of the tether, the status of tether entanglement, and a running count of the number of times the tether has collided with an object.

The ROV can be controlled either with a joystick or the keyboard. The intended interface is the joystick because it more closely simulates the operating environment for remote control of an underwater vehicle. The joystick used for this project is the Microsoft Sidewinder 3D Pro. This particular joystick was chosen because, in addition to the two degrees of freedom that a traditional joystick possesses, it also has a third degree of freedom called rudder control. Rotating the joystick handle operates the rudder control and controls the rotation of the ROV.

The trainer's ROV dynamics were loosely modeled after the Imetrix ROV, the Talon. Talon has four bidirectional thrusters mounted at approximately 45-degree angles at each of the four corners of the ROV. Four motors control the horizontal movement and rotation of the ROV. A fifth motor is positioned vertically in the center of the ROV and controls vertical movement.

The trainer has the ability to record a simulation scenario to a log file or to playback log files of previously recorded simulation scenarios. The part-task trainer requires a minimum configuration of a 60 MHz Pentium computer or equivalent running Microsoft Windows 95 or Windows NT 4.0. The program was developed in C++ using the Microsoft Developers Studio and uses version 4.2 of the Microsoft Foundation Classes (MFC) libraries.

Integration of Tether Model into General Trainer

The collision detection code currently used in the general trainer only works for detecting collisions between the front of the ROV and objects in the environment. A better collision detection system is needed for detecting collisions in which the tether is involved. We have selected the VCOLLIDE system from the University of North Carolina to form the basis of the new collision detection system. This system, which is built upon the RAPID collision detection system developed by the same group, is an accurate and fast collision detection library which uses triangle-based models and object-aligned bounding boxes for its collision engine. However, for use in this project, several modifications have been made related to the tether modeling. The unmodified system will report when pairs of objects collide, but give no other information about the collision. The main modification to the system enables it to also report which triangles in the models are involved in the collision. With this information, an approximate point of contact between the tether and other objects is relatively easy to calculate. As a final addition, a wrapper class for easily adding Open Inventor objects to the VCOLLIDE database has been created.

Although we have not completed integrating the tether model into the general training system, certain preparatory tasks have been accomplished. The straight-line-segment and the V-shaped models have been extended to 3-D, and the collision detection procedure has been modified to take account of this transition. Whereas in the 2-D models, each object in the scene is specified along with certain bounding points at which the tether can make contact, in the 3-D models, no such collision points are specified. Thus, the modified VCOLLIDE package is used to determine these collision points. The triangles in each of the models that are involved in the collision are used to determine an approximate point of collision, and, as in the 2-D system, this collision point is used, along with the relevant angles, to determine how the tether reacts to the collision (e.g., whether the tether breaks up into two segments at the point of collision).

1.7.3 Sonar Simulation

In a separate project, we collaborated with other members of the TRANSoM team at BBN and Imetrix to provide the general trainer with a sonar simulation. This simulation, the physical modeling for which was done primarily by other members of the team, is intended to resemble the Imagenix Model 881 scanning sonar.

In each of certain scanning positions, the model sends out several rays in an arc from the vehicle location until either a maximum range or an object is reached. When an object is detected, it generates a return with an intensity based on the angle between the incident ray and the surface of the object. All objects are assumed to have the same sonar reflectivity. Each echo is recorded, along with its position, intensity, and generating object. The sonar then moves on to the next scanning position until a desired sector width has been scanned. Then the sonar scans back in the other direction. The training angle (angle from straight-ahead heading) and sector width of the sonar can be set as needed.

The sonar display that we have developed has three main components. First, the main frame displays the sonar returns, range arcs, and bearing lines. When the sector width of the sonar is adjusted, the sonar display resizes to make the sector as large as possible while maintaining a true 1:1 aspect ratio. Second, the sonar controls allow the user to adjust the range, sector width, and training angle of the sonar. The range of the sonar can be set to 5, 10, 25, 50, or 100 meters. The sector width can be set to 15, 30, 45, 60, 90, 180, and 360 degrees. The training angle can be changed in three-degree increments. Third, the status component contains a display that shows the area scanned by the sonar as a filled arc in the correct direction relative to the vehicle.

In the real sonar system, the user can click on one position and drag the mouse to get a differential range and bearing between the current and selected point. As currently designed, this feature is not implemented in the simulated sonar; however, it is used for the situational awareness map display.

1.7.4 Auditory Display of Thruster State

This section describes work directed toward exploring ways in which “continuous” information, particularly information relevant to the state of the ROV thrusters, can be displayed as a spatialized sound field. Such a display constitutes a separate channel of information that is independent of the “discrete” (speech-based) channel presently used to provide verbal tutoring. By providing continuous representation of thruster state, situation awareness within the training session can be enhanced, thereby instilling not only a redundant sense of the velocity vector of

the ROV, but also a heightened awareness of the interaction of thruster power, thruster condition, water current, and tether drag. Such enhanced situation awareness could transfer to real-world operation of the ROV even in the absence of the spatialized sound field.

There are three components to this effort: (1) creation of an appropriate auditory testbed for generation and spatialization of auditory stimuli; (2) development of plausible display formats to be compared and evaluated; (3) carrying out of a series of formal experiments to objectively evaluate the proposed display techniques.

Development of Audio Testbed

The audio testbed includes equipment for generating six audio source channels having independently controllable waveform parameters (amplitude, frequency, timbre, etc.); these sources are then spatialized through the use of a computer-controlled convolution engine allowing filtering, delays, and amplitude control to be applied to the signal as needed. Much of the ancillary equipment, such as loudspeakers, headphones, amplifiers, and head trackers necessary to produce the desired acoustic field from the electrical representation of these stimulus signals was already present in our laboratory. This existing equipment was augmented with additional items purchased as part of our DURIP-funded PC VE testbed upgrade.

Specific major equipment items comprising the Audio Testbed include a single PC running three SoundBlaster cards yielding a total of six channels, a Convolvotron, and an Ascension head tracker. A set of six loudspeakers with associated amplifier and control electronics has been added to support off-head display studies in addition to headphone-based presentation. (Recently, a more advanced Tucker-Davis system which would subsume the SoundBlaster and Convolvotron functions has become available to us due to the completion of another project in our lab. We are modifying our software to take advantage of this improved instrumentation.) The current system is configured to operate under Windows NT, and software has been written to implement two alternative display approaches, *virtual thruster* and *simultaneous thrusters*. In the case of the virtual thruster, we are able to image a source (one channel of a single SoundBlaster card) using the Convolvotron; in the simultaneous thrusters display we use the multiple outputs of the SoundBlaster cards mixed (with simple filtering and amplitude tweaking) into the multiple audio channels to create the test stimulus.

Approaches to Audio Spatialization

The display formats that we are studying center on two general approaches. The first approach is to create an abstract representation of a single *virtual thruster* which is spatially imaged at a fixed position relative to the pilot's head. In this representation, the angular position of the image corresponds to the ROV direction of motion, and the intensity and timbre of the thruster sound provides an indication of the magnitude of the ROV velocity vector. The second approach involves creating *simultaneous thruster* sound sources placed at fixed positions about the pilot's head to create a sense of presence within the ROV. In this approach, the relative sound qualities serve to convey the ROV velocity vector.

Although the first approach to displaying this information appears to be more straightforward for the purpose of indicating the ROV velocity vector, it is not clear that it serves other aspects of the situational awareness goals. Moreover, due to the “low resolution” of both auditory localization (from a psychophysical standpoint) and auditory spatialization technology (due to the need for acquiring individual HRTFs to enable really precise spatialization), it would be more efficient to simply include a visual indication of the ROV velocity vector on the video display if this were the major consideration. To resolve the question of the utility of displaying this information in audio versus visual modalities, it will be necessary to carry out a set of experiments in which performance using the alternate systems is characterized along the two domains involving (1) the ROV velocity vector information, and (2) ROV state information (a component of situational awareness).

Mapping of System State to Audio Domain

It is not obvious that merely providing an indication of the velocity vector through the audio channel is a compelling reason to increase the system complexity, especially when this information could be easily incorporated into the existing visual display. However, consideration of the importance of the audio channel in manned vehicles, where sounds (especially from the engine) are frequently available to the pilot, we have developed a system to augment the audio display to reflect the operational state of the thrusters. By including this information in the ROV

simulation, we expect to enhance situational awareness in a manner that may instill insight to the pilot that would extend even to cases where this additional information is not present.

In a real-world thruster, the correlation between the RPM of the screw and the power sent to the thruster motor is not always constant. Deviations from the typical correlation are due to the varying (environmental) conditions in which the thruster is operated. For example, if a strong current is flowing along the axis of one thruster, that unit will require more power to maintain a given RPM if the current is opposing the thruster and will require less power if the current is in the same direction. In this case, the changes in the relation between RPM and power would convey information about the water current. Similarly, various other loading factors (tether drag, contact with obstacles) and fault factors (seaweed entanglement, loss of mounting integrity) will change the RPM/motor power relation in characteristic ways.

The display method we have devised takes advantage of the psychophysical separability of pitch and loudness to convey the electromechanical relation in a very natural "ergonomic" way. Using this technique, the factors affecting thruster RPM (basically, all the loading and fault factors listed above) are used to determine the pitch of the sound representing the thruster while the power supplied to the thruster motor is used to determine the loudness (amplitude) of the presented sound. This mapping is applicable to both the virtual thruster and the simultaneous thruster display methods as long as the loads and forces are calculated with respect to the 3-D orientation of these vector quantities. This computational housekeeping is trivial to perform since all of the required information is available from the simulation process.

At present, these algorithms are implemented within a simple test program in which each input can be "dialed in" and the result heard on the audio display; incorporation into the simulation for experimental evaluation should be straightforward.

Integration and Evaluation Issues

Although simple comparison tests (between various display methods) can be performed in a part-task trainer environment, situational awareness issues involving thruster state are likely to become clearer if testing is performed in the full Imetrix trainer. Incorporating these elements into the trainer for the purpose of performing these evaluations should not be difficult, because the audio display routines have access to the necessary variables (joystick position, faults, currents, tether load) from the ROV and tether simulation processes, and because there would be very low computational overhead in servicing the audio display routines.

Additional factors to be considered in evaluating the feasibility of adding audio displays of the type described here include incremental equipment cost (e.g., spatialized audio over headphones requires significant additional equipment) and the space requirements of a field-deployable training system. In general, multipoint displays (i.e., simultaneous thruster) requiring separate speakers do not require expensive head tracking equipment. However, separate speakers may not be practical in field deployment due to ambient noise and space concerns (e.g., a set of loudspeakers located behind the pilot may interfere with other space requirements in cramped locations).

The alternative single point display methods (i.e., virtual thruster), whether they use headphones or off-head loudspeakers, require headtracking. However, if a headphone display is required, single point display methods entail a smaller cost increment because the head tracking equipment is required anyway. Therefore, selection of the best approach to enhancing the display of information through the audio channel will require specific knowledge of the relative merits of the displays considered. We hope to be able to provide this information through future experiments utilizing the technology assembled during this past year.

1.7.5 Force-feedback, Training-transfer Experiment

Introduction

In these experiments (which constitute a continuation of experiments performed and reported on earlier in the program) a human test subject controlled a simulated underwater remotely operated vehicle (ROV) by applying simulated thrust via a joystick. The joystick was capable of providing various types of force feedback as specified by the experimenter. In the normal or "Control" condition of the experiments, force on the subject's hand was proportional to the thrust (like a common spring-return joystick). The goal of the experiments was to see if some type of additional force feedback during training might

enhance the subject's ultimate performance using the control condition. In all cases, the movement and force feedback was constrained to one dimension.

The training strategy used was to maintain the normal force/position relationship of the joystick in a small region about some optimal control position. When the subject's control input deviated too far from the optimum point, the spring constant of the joystick increased abruptly, giving the subject a haptic sensation that he/she was out of bounds and physically pushing his/her hand back toward the desired region. In this sense, the strategy is similar to that used in a pair of training wheels on a bicycle, allowing the subject to feel the normal force/position relationship as long as he/she stays within the desired region. The optimum position varied with time depending on the task and the performance metric that was used to define “optimum.”

The main result of this experiment was similar to that of the one reported previously: that the additional force feedback did not enhance training.

Development

Developmental work was required in two areas in order to carry out the experiment: (1) establishment of a performance metric suitable for analyzing the test results and (2) design of an optimal controller for use as a reference in training and for analyzing the results.

The experimental task was defined so that it was effectively the same in each trial, with the subject starting over each time. A discrete tracking task was employed in which the subject was required to respond to step changes in the target position. The performance metric used with this task was the mean squared position error between the position of the subject-controlled ROV and the target position. The error was measured every 1/60th of a second (i.e., each time step of the control loop) and summed over a given trial period.

Our analysis suggested that the mean squared error would be minimized in a step task by the following control input. First, at the start of the trial (when the target changes to a new position) the ROV thrust should immediately be set to maximum (pushing the ROV toward the target). Full thrust is maintained until, at just the right moment before reaching the target, full reverse thrust is applied to the ROV. Full reverse thrust is maintained until the ROV comes to a stop exactly at the target position, at which point the thrust is set to zero.

Although a closed-form solution was derived for this *optimal bang-bang controller*, we realized that it would not be entirely adequate for the experiments because: (1) it would result in limit cycles about the set-point if actually implemented as a discrete-time controller; and consequently (2) it did not provide an adequate model for training the subject. What was needed was a controller where the control input is reduced in some optimal way as the set-point (ROV target) is approached.

Our solution was to design a PID controller by trial and error so that it gave adequate step response (not too different from the bang-bang controller) for step position changes ranging from 0.1 to 10 meters. Thus, while not optimal in the sense of minimizing the mean squared error, the PID controller approached the performance of the optimal bang-bang controller in step response. It also gave a good response to a 0.1 hertz, 1-meter-amplitude sine input.

The controller included a 0.15 second pure time delay to allow for the neuromuscular lag of the human subject. This was done so that it would be possible for the subject to keep up with the training feedback.

Methods

Subjects with only 10-15 minutes of written and verbal instruction (and no manual practice) attempted to control a simulated ROV represented by a green cross on a computer monitor. The task was to minimize the position error between the green cross and a target (represented by a red line) that periodically moved in discrete random jumps or position steps. The motion of both the ROV and the target were constrained to motion along a horizontal line.

To control the ROV (green cross), subjects manipulated a joystick. Pushing the joystick to the right applied a simulated thrust causing the ROV to accelerate to the right according to a simple dynamic model that incorporated linear damping. The greater the displacement of the joystick, the greater the applied force. Similarly, pushing the joystick to the left applied thrust to the left, and zero simulated thrust was applied to the ROV when the joystick was at its middle position.

The subjects performed the task during two 16 minute trial periods, separated by a break of approximately five minutes. The first period was a “training” period, and the second was a “test” period. The subjects were divided into two approximately equal groups that were trained in different ways. The two groups of subjects received different types of force feedback from the joystick during the training period. Then both groups performed an identical “test” task (with the same type of force feedback) during the second period.

During the second trial period the force feedback of the joystick was like a conventional spring-return joystick, with the force applied to the subject's hand proportional to the displacement of the joystick. The “Control” subject group was trained with the same type of feedback during the first trial period, while the “Experimental” group received the additional force feedback described previously. The joystick algorithm included a dead band region at the center position, with barely perceptible discrete changes in the force feedback at the edges so that the subjects could feel when the joystick was in the neutral position.

Twenty people from the MIT community (primarily undergraduate students, roughly half male and half female) served as subjects in the experiment. They ranged in age from 19 to 45 years. Nine subjects were in the control group and 11 subjects in the experimental group. All but two of the subjects were right-handed, and all considered themselves to have normal vision (corrected in some cases) and normal hand-eye coordination. Each subject performed the test only one time. They had no prior experience with the experimental task, and they were paid for their participation.

The experiment was performed using the Virtual Workbench. An article on this facility has been submitted for publication. The key features of the apparatus as used in the present experiment were as follows. First, the 3 degrees-of-freedom (dof) force feedback PHANToM™ manipulandum served as the joystick which the subject used to input thrust commands to the simulated ROV. Second, the green cross representing the ROV and the red line that served as the target were presented on the computer monitor. Aside from the fact that the monitor was viewed in a mirror, the display was an ordinary computer monitor with the ROV and target appearing as two-dimensional images on the surface of the screen. The PC incorporated in the apparatus generated the images on the display, controlled the PHANToM™, and recorded the data.

A 3/8-inch thick acrylic plate was mounted on the frame of the virtual workbench (VW) to constrain the motion of the PHANToM™. The plate was mounted horizontally at approximately the level of the subject's hand. The standard 3 dof gimbal was removed from the end of the PHANToM™ and a rigid thimble fixture was substituted at the end of the last link. The link (a rod approximately 0.25 inches in diameter by 6 inches long) passed up through a slot in the plate such that it was free to move from side to side. In the experiment, the subject held the thimble fixture at the end of the link and moved the link back and forth as if it were a large joystick. Although the link moved freely in all directions, the slot/plate only allowed large motions in the left-right direction.

Like the joystick, the graphical images on the display were constrained to a horizontal line (x direction, running left to right) across the center of the display screen. The target position (x coordinate) was specified by an “action” file. The program that was running the test read the action file (a text file) specified by the experimenter at the start of the test. The trial ended when the computer read a special code at the end of the file. A piece of paper was placed under the partially silvered mirror so that the subjects could not see their hands when looking at the display in the mirror.

The experimental task consisted of ten blocks of twelve standard steps for a total of 120 target position changes. The standard steps were randomly ordered from block to block, but summed to zero so that the target returned to its initial position in the center of the display at the end of each block of steps. The target changed position every eight seconds, allowing the subject plenty of time to reach the target before the next step. Note that the first step occurred three seconds after the start of the test, but in this case the initial positions of the ROV and the target were the same.

The performance of the experimental group was initially worse than that of the control group during the first block of 12 steps. Then, by the end of the first trial period (consisting of ten blocks of twelve steps), the performance of the experimental group was slightly better than that of the control group. Thus, it appears that the additional force feedback was initially confusing to the subjects but later provided an

additional useful cue. The performance of both groups was approximately the same during the second trial period when both groups performed the task without additional force feedback (like the first trial period of the control group). The control group may have done slightly better than the experimental group during the second period, but the difference was small in comparison to the fluctuation in the performance in the mid-to-latter part of the period. Possibly some of the subjects became tired and/or bored.

No statistical tests were performed on the data. However, independent of any small effects that might be shown to be statistically significant through such analysis, it is clear that the additional force feedback had no important effect on ultimate performance during the training period nor on the subsequent performance during the test period (i.e., there was no positive training transfer).

1.7.6 Theses

Esch, M. *Determining the Position of Underwater Tethers in Real-Time.* B.S.E.E./M.Eng. thesis, Department of Electrical Engineering and Computer Science, MIT, 1998.

Zalesky, J.L. *A Part-Task Trainer for Underwater Tether Awareness*. B.S./M.Eng. thesis. Department of Electrical Engineering and Computer Science, MIT, 1998.

1.8 Training Spatial Knowledge Acquisition using Virtual Environments

Sponsor

U.S. Navy - Office of Naval Research
Grant N00014-96-1-0937

Project Staff

Nathaniel I. Durlach, Dr. Thomas E.v. Wiegand, Rebecca L. Garnett, Xudong Tang, Andrew G. Brooks, Samuel R. Madden

1.8.1 Introduction

A general overview of this project is presented in *RLE Progress Report No. 140*.[36] During the past year, our work has focused on (a) development of a locomotion interface device, (b) development of a virtual environment (VE) construction system, and (c) preparation of reports.

1.8.2 Locomotion Interface Device

One important aspect of a VE training program designed to facilitate the acquisition of knowledge about a space involves providing the user with a plausible means of moving about within the virtual environment. Our research during the past year has included the development and testing of an inexpensive interface that allows a user to “finger walk,” or simulate, by means of finger motion, the activity of walking within and through a virtual environment.[37] While there seems to be some direct relationship between the development of spatial knowledge and the amount and type of effort expended in moving from place to place (as one expends effort while walking in the real world), the choice of a motion-control interface for exploration within virtual environments may be constrained by factors such as cost. The interface which we have designed is one which operates within the well-known “walking metaphor” of motion control, making use of a low-friction pad that allows the user to “walk in place” by means of moving his fingers, and an electric field sensing system that monitors the position of the fingers on the pad. The interface effectively tracks the user’s movement along the surface of the pad for input into the virtual environment.

The potential benefits of this work are twofold. First, it is possible that many of the expected advantages of a full-scale walking interface can be realized in a more cost-effective manner by means of a scaled-down, finger-walking interface. Second, the experience gained in developing such a finger-walking interface using a slippery pad may be useful for subsequent work on a slippery-floor walking interface.

The finger walking device is an inexpensive, compact, easy-to-use interface for providing locomotion within virtual environments. The operator uses a nat-

36 N.I. Durlach et al., “Training Spatial Knowledge Acquisition using Virtual Environments,” *RLE Progress Report No. 140* (Cambridge, Massachusetts: MIT Research Laboratory of Electronics, 1998), pp. 398-99.

37 S.B. Fitch, *The Finger Walker: A Method to Navigate Virtual Environments*, M.Eng. thesis, MIT Department of Electrical Engineering and Computer Science, 1998.

ural walking-like motion with fore and middle fingers, with minimal equipment attached to his body. The input to the user interface is a tracking of the change in the electric field created by the user's fingers. The output to the virtual environment from the finger walker is a velocity vector, consisting of a magnitude and a direction.

The operation of the interface is easy and straightforward. First, the user sits down at the computer, workstation or other setup for viewing the virtual environment. The user then attaches transmitter electrodes to his or her fingers for tracking. Next, the user places an HMD on his or her head or positions him or herself before a standard computer monitor, to view the virtual environment. Finally, the user places his or her fingers on the finger walker pad and begins moving his or her fingers in a walking-like motion. The finger walker and the virtual environment software perform the calculations which update the position of the user in the virtual environment. The user interface consists of five distinct stages of operation: signal detection, data acquisition, translation, special operation instructions, and velocity computation. Hardware systems detect the electric field and send the data to a computer for processing. The finger walker software package then manipulates the electric potential received from an analog to digital card to compute a velocity vector.

During operation, the coordinates of the user's fingers and the velocity of the user through the virtual environment are displayed within an on-screen window, which also provides a graphical positions map and a directional compass. A separate window tracks the movement of the user through the virtual environment by drawing a line along the path of the user as determined by the magnitude and direction variables.

Initial tests of the interface have provided evidence to support the electric field proximity sensor as an efficient method by which to track the movement of the user. The finger walker appears to be fairly accurate when tracking the position of the fingers across the pad, and the tracking window shows that the device provides an effective means of moving through a virtual environment. Some limitations of the present device include a lack of memory of past system events, and extreme sensitivity of the hardware to slight changes in the position of the finger, as well as to noise in the system. It is expected that improvements to the system can be introduced to eliminate these difficulties and that the effectiveness of the "walking" motion and expenditure of effort on the user's ability to estimate distances accurately in the virtual environment should then be subject to future experimental study.

1.8.3 The Virtual Environment Construction System

The most time-consuming task in the creation of any VE simulation is the construction of the 3-D models and the acquisition of photo-realistic textures for these models. The VE construction system, which we have nearly completed this past funding period, simplifies both of these tasks by allowing users to import two-dimensional DXF floorplan files (commonly available for many installations, including all of the buildings at MIT) and then to automatically attach textures to walls, doors, windows, and other objects within the building.

Three-dimensional open-ended architectural database system (TOADS) is the software component of this system. TOADS can import plan files of various types, due to the open-ended design of the program. The indoor worlds of our current work are based on DXF files available from the MIT space accounting office. Using the graphical interface tools provided within TOADS, we can group these raw floorplans, which initially consist of uncoordinated lines and polygons, into logical objects that are classified by object type (e.g., room, door, wall.) A graphical interface allows floorplans to be further augmented with paths indicating the areas through which VE subjects will normally walk and with movable polygonal objects that represent non-permanent structures (such as furniture) within the building.

The hardware component of the system supports the texture acquisition function. It consists of a rack-mounted scanner camera which runs on a linear path. A software controller (the "Grabber" program) steps the camera along a section of wall, capturing image frames at regular intervals, and pastes them together to form long strips of texture for each wall section. These texture files are automatically managed by the software and are associated with specific polygons within the floorplan.

Once textures have been associated with the floorplan, the datafiles can be exported as a list of 3-D polygons and associated textures in any format required by the real-time VE rendering software. Currently, we are using VRML-format output from the

system for use with the Cosmo Player VRML browser (non-immersive viewing) and DIVE (for immersive HMD-based viewing).

Hardware Development

During this year, we have completed our work on the rack-based texture scanning system begun last year. During this work, we have solved a number of problems which previously degraded the captured images, and we have refined the controlling interface to be more closely integrated with the TOADS software (discussed in detail below). The present scanning system utilizes a Macintosh G3 laptop computer with a video capture card for image capture and general interfacing and control functions, while the control of the stepper motor and position sensing along the rack are controlled by a “BASIC Stamp” microcontroller via a serial port.

Our work on the scanning rack has resulted in two slightly different versions. Our main rack utilizes a 5.5 foot toothed rack and a mating pinion gear on the stepper motor. The camera and motor are mounted on the moving carriage. The second rack is a 4 foot unit, belt driven, in which the motor is stationary; only the camera (and an optional microphone) is mounted on the moving carriage. Both racks utilize identical electronics to control the motor with only a slight change in the firmware to accommodate the different rack lengths. Both racks include similar provisions for making initial adjustments to the tilt and rotation of the camera, as well as a simple crossed-beam laser alignment array.

One departure we have made from earlier scanning systems is to include a video-capture card and video camera as a replacement for the QuickCam. Although the QuickCam was usable for some purposes, it was not suitable for experimenting with different (pre-distorting) lenses. In addition, the resolution was about half that of even an NTSC camera, and the compression required to send the QuickCam output over the serial port was seen as an unnecessary constraint. The current video-capture system takes advantage of the special hardware characteristics of the G3 PowerMac, allowing fast and flexible video processing, with our choice of video camera. Rack 2 utilizes a fairly standard 1CCD color NTSC camera fitted with a fisheye lens, Rack 1 contains a more elaborate Y/C camera which contains a dataport for controlling camera parameters (white balance, exposure, etc.).

The purpose of using the fisheye lens is to achieve better utilization of the available pixels within the field of view. In this way, a larger percentage of pixels represent the world at and near eye level, and fewer pixels are dedicated to the periphery (floor and ceiling areas). This compression at the periphery is compensated for within the capture program, so that objects maintain their natural proportions regardless of where they appear in the field of view. Only the image resolution changes (decreases) as the object moves to the periphery.

Lastly, a provision for generating synchronized acoustic impulses has been included in the form of an instrumented cap pistol. Although the report of the pistol is not as distinctively impulsive as an electronically generated sound (e.g., a spark or piezo generator), the output is very high, and the average shape of the waveform is predictable enough to determine the room impulse response if the known waveform of the pistol is taken into account.

Software Development

The parsing engine is designed primarily to work with floorplans provided by the MIT space accounting office (http://web.mit.edu/ofms-space/www/). The parser is limited in the types of DXF objects it supports. In particular, it lacks support for rotated objects, for hatched and patterned objects, and for blocks (object groups). The parser has not been tested with files other than MIT blueprints and may or may not behave properly when non-MIT files are used.

Selecting “Open…” from the File menu will bring up a standard file dialog, from which a text file can be selected. If an invalid or uninterpretable DXF file is selected, nothing will happen. Otherwise, a window showing a building blueprint will open. A tool window will also appear. Once a file has been opened, it can be manipulated via any of the tools shown in Figure 5. Navigating the file is accomplished via the zoom tool and the scroll bars. The zoom tool doubles the size of drawn objects and tries to center the picture at the clicked point. If the option key is held down, objects are made half as large (the document is zoomed out).

Objects can be selected via the Select Tool or the Select Box Tool. The Select Tool selects any objects beneath the clicked point on the screen. The Select Box Tool selects any objects completely enclosed in the dragged rectangle. Selected objects can be removed by hitting the delete key.

Table 5: Description of Tools.

No.	Tool Name	Description
1	Zoom Tool	Zooms into or out from (by holding down the option key) the selected point in the document.
2	Room Tool	Defines a room as a closed polygon of points. Brings up the Room Creation dialog.
3	Select Tool	Selects objects under the selected point in the document.
4	Grabber Tool	Creates the 5′x5′ texture grabber box, which can be dragged along lines in the document and used to associate the texture grabber dialog with a specific piece of an object.
5	Junk Tool	Creates polygonal objects in the document which represent movable, non-floorplan specified items (such as furniture, boxes, and trash) in the environment.
6	Path Tool	Creates parallelogram-segment based paths. Paths have two collision detecting walls and two non-collision detecting walls. They define habit-trails where people can move freely through junk.
7	Box Select Tool	Allows objects within a specified rectangle to be selected.

Several menu items become available when an object is selected. The DXFObject menu controls settings specific to individual objects. The “Set Object Length...” item allows a correspondence to be established between DXF coordinates (as arbitrarily defined in the DXF file) and real world distances (as measured in the real building). This setting is only important in determining the size of the grabber rectangle. Initially, 1 DXF unit is assumed to be equal to 1 meter in length. Selecting “Get Object Info…” brings up the Object Info Dialog for the first-selected item. The Object Info Dialog allows a PICT file-based texture to be associated with an object, allows an object to be given a specific height, and allows collision-detection to be turned on or off for the object. The “Collision Detect” item allows the collision detection flag to be turned on or off for any number of selected items. A checkmark will appear next to this item if the first selected object has collision detection turned on. The “Set Object Type” item allows a group of objects to be set to a specific object type. Object type affects the default texture associated with an object, as set through the preferences dialog (see the next subsection). Currently supported object types include walls, doors, windows, elevators, stairs, junk, and paths.

In addition to object specific settings, there are a few document-level settings and defaults that are accessible through the File menu under “DXF Preferences…” Default textures can be set for each of the object types specified above. A default ceiling height for objects which specify no height can also be set. Finally, the mapping between real-world and DXF-space coordinates can be set, as in the Object Length dialog described above.

The Junk and Room tools have similar interfaces. They define closed polygons on the floorplan which correspond to the outline of junk or the boundaries of a room. The tools work much like the polygon tools from common drawing programs: each mouse click on the screen defines a point in the polygon. A double-click closes the polygon.

When the room tool is in use, the Room Settings dialog will appear after the polygon has been closed. This dialog can be used to specify a name, color, ceiling height, and texture for a room. The “Rooms” menu can be used to add objects to rooms or change the settings for a room. Rooms are a logical unit that contain objects; objects inherit height and texture information from their parent rooms. Once a Room or Junk polygon has been defined, it can be dragged freely around the drawing and rotated about its center. This is particularly useful for junk, which may be moved in the physical building and need to be updated in the floorplan.

The rooms menu allows objects to be added to rooms, rooms to be selected, and the settings for rooms to be changed. Selecting a room from the “Add Item(s) to Room” submenu will cause the cur-

rently selected objects to become members of the specified room. Objects can belong to several rooms, because there may be walls which are on the boundary between two logical room areas, although objects only inherit height and texture information from the first room to which they are added. There is currently no way to remove an object from a room once it has been added.

Textures are added via the Grabber tool. When the Grabber tool is initially used, it places a 5 foot-by-5 foot rectangular box on the drawing at the mouse location. This box represents the texture acquisition frame and can be dragged around the drawing. It will automatically snap to the nearest line object in the drawing, or to the center of the nearest junk polygon. It rotates to match the slope of the line it is currently snapped onto. Pressing the left and right arrow keys with the Grabber tool selected will cause the Grabber rectangle to move parallel to the line it is currently snapped to, in increments of the rack length.

Once the Grabber tool has been snapped to an object, the “Grab Texture…” item in the DXFObjects menu becomes enabled. Selecting this item brings up the Texture Acquisition dialog. A scrolling list shows QuickTime movies which have been scanned using the Grabber Program or added from an external source. Clicking the “Add” button brings up a standard file dialog which can be used to select any QuickTime movie. The “Scan” button launches the Grabber program and commands it to acquire a texture-movie.

A second scrolling list contains PICT files which represent swaths of texture. They are obtained by tiling frames of captured texture movies into a single file. Clicking on a movie in the left list and pressing the “Copy >>” button launches the QuickTime Viewer application and opens the selected movie. QuickTime Viewer processes the movie to create texture swaths, which are sent back to the dialog and added to the “Processed Textures” list. “Up” and “Down” buttons can be used to arrange pictures so that textures corresponding to adjacent wall segments are next to each other in the list. Once textures are properly arranged, clicking the “Make” button concatenates them into a single texture file.

The Grabber program is designed to capture QuickTime movies via QuickTime compliant FrameGrabber devices such as the Connectix QuickCam. When the Grabber program launches, it attempts to open the first available frame grabber device. If it finds no such device, it exits. Otherwise, a small window with the input of the current device appears. To change settings for the current frame grabber, or switch to a new frame grabber, select “Camera Settings…” from the Video menu. This will bring up a QuickTime Input Settings dialog box, which includes a “Source” panel, used to select the current input device.

The Grabber application is designed to control a simple stepper motor via the Mac serial port. It drives a Basic Stamp based controller via a simple serial interface. Selecting “Robot Settings…” from the Robot menu brings up a dialog which can be used to select the serial port for the motor controller and the size of the steps which the motor makes. The Grabber program can also capture sound impulses as it records video. The “Impulse Settings…” item in the Sound menu shows a dialog which allows specification of the duration of the sound impulses associated with video frames as capture occurs. Impulses apply to all video frames from the time they are captured until the time the next impulse is captured.

Once the settings have been properly configured, movie capture can begin. When “Start Recording…” is selected from the File menu, a standard file dialog will appear requesting the location of the recorded movie. Then, the QuickTime Input Settings dialog will appear. Go to the “Compression” panel and select the desired compression method. Exit the dialog by clicking “OK” to begin recording. For maximum recording efficiency, events are not processed and the main video window is not updated. To add an audio impulse during recording, type ⌘-I. The sound input device in the sound control panel will record for the specified period of time (no visual cue will be given). To stop recording, type ⌘-E. The movie will be written to disk and the program will return to normal operation.

The QuickTime Viewer application allows examination of all data stored in the augmented QuickTime movies generated by the Grabber program. It takes these movies and concatenates rectangular sections of frames together to form panorama-style PICT images. When a Grabber-generated movie is opened in QuickTime Viewer, a list of frames in the movie is presented. Double-clicking on a frame displays the frame and some basic information about it.

Panoramas can be generated by highlighting the frames which are to be included and selecting “Build Panorama…” from the File Menu. Frames can be clipped to a rectangle using the “Select Frame

Size…" option from the Panorama menu. Once the clipping is set up properly, select "Save To PICT" from the Panorama menu to save the texture to disk. The "Compute Frame Size…" option in the Panorama menu shows a dialog specifying the wall-to-camera distance, camera field-of-view, and camera movement per frame. The program will automatically compute the appropriate frame size to exactly tile the movie. Finally, the Panorama menu includes the items "Rotate Clockwise" and "Rotate Counterclockwise"—these options rotate each frame 90° in the specified direction, to compensate for a horizontally oriented camera (which allows more vertical resolution.)

The TOADS program does not directly support any existing 3-D file format but instead generates an intermediate file format which contains information about the 3-D polygons and textures which should be included in any 3-D file. The file format and the C++ interface to it are available in the on-line TOADS documentation. The details of this process are transparent to the user once the appropriate output module is written and installed.

The current output module from TOAD generates VRML files that we are using in conjunction with DIVE real-time immersive VE rendering software. The DIVE package runs both on the SGI platform and on Windows machines. However, we are primarily using the DIVE/WindowsNT/Intergraph platform that has recently been assembled in our lab. DIVE is the product of academic research and is not part of a closed-environment for-profit enterprise. We can get source code to modify the software to our needs, and there is a community of users to help with implementation issues. DIVE is available for both Unix and Windows/NT platforms, and indeed we have it running on our SGI machine as well as the Intergraph.

DIVE also comes pre-configured to support collaborative VEs with multicast sound and video. The ability to create "invisible participants" and the ability to record user responses and activity (motions, viewpoints, actions) are important features for use in an experimental environment. Furthermore, many specialized features have been added (and are available at no extra cost) by the users, such as spatialized audio support.

World descriptions in DIVE are VRML-based, further enabling crossplatform sharing and development. VRML browsers are available for all platforms, and VRML has been accepted as the current standard for VE design and distribution. Creating our environments in VRML means that we will be able to extend them using environments others have built and distribute them without the inclusion of bulky, platform-dependent executables. DIVE also includes an API whereby many of the limitations of VRML can be overcome via C and tkl/tk programming. This API enables many of the multi-user and multicast features of DIVE and also allows highly interactive, dynamic environments, which are not well supported by raw VRML.

Preparation of Reports

The experimental results obtained by Koh[38] on the acquisition of configurational knowledge have been further analyzed and prepared for publication.[39] In addition, work has been initiated on the development of a "white paper" for use by the Office of Naval Research in planning research for this field.

Journal Article

Koh, G., T.E.v. Wiegand, B. Shinn-Cunningham, R.L. Garnett, and N.I. Durlach. "Use of Virtual Environments for Acquiring Configurational Knowledge about Specific Real-World Spaces: I. Preliminary Experiment." *Presence: Teleop. Virtual Environ.* Forthcoming.

Thesis

Fitch, S.B. *The Finger Walker: A Method to Navigate Virtual Environments*. M.Eng. thesis. Department of Electrical Engineering and Computer Science, MIT, 1998.

1.9 Further Results on Supernormal Auditory Localization

Sponsor

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

38 G. Koh, "Training Spatial Knowledge Acquisition using Virtual Environments," M.Eng. thesis, Department of Electrical Engineering and Computer Science, MIT, 1997.

39 G. Koh, T.E.v. Wiegand, B. Shinn-Cunningham, R.L. Garnett, and N.I. Durlach, "Use of Virtual Environments for Acquiring Configurational Knowledge about Specific Real-World Spaces: I. Preliminary Experiment," *Presence: Teleop. Virt. Environ.*, forthcoming.

Project Staff

Nathaniel I. Durlach, Professor Richard M. Held, Dr. Barbara G. Shinn-Cunningham, Douglas S. Brungart, Salim F. Kassem

1.9.1 Overview

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

1.9.2 Recent Work

During the past year, we have both prepared many of our previous results for publication[40] and completed a new study on adaptation to an enlarged head in which the simulation of the enlarged head is achieved by frequency scaling the head-related transfer functions.[41]

In the new study, two experiments were performed. Both experiments used a forced-choice identification task with 13 different azimuthal positions. In the first experiment, which included 40 runs in each of eight sessions for each subject, the first two and the last eight runs used normal cues, while the middle 30 runs used the altered cues. In this experiment, correct-answer feedback was provided after each response. In the second experiment, which employed only ten runs per session per subject, the first two and the last runs used normal cues, while the middle seven presented altered cues. In this experiment, feedback was not provided. The experiments showed that trial-by-trial feedback accelerates the adaptation process to supernormal cues, but that it is not necessary for adaptation to occur. For both experiments, mean response and bias showed all the usual characteristics of adaptation. The resolution results, however, were less consistent. Although changes in resolution in the feedback experiment were similar to the changes seen in our previous experiments, without feedback the postulated internal noise was large throughout the experiment. Subjects seemed to attend to the whole range of possible cues encountered in the whole experiment, independent of the adaptation rate. In general, resolution was better at the center positions when the altered cues were introduced, but normal cues provided better overall resolution.

These results have not yet been fully analyzed or prepared for publication.

Journal Articles

Brungart, D.S., and W.M. Rabinowitz. “Auditory Localization of Nearby Sources I: Near-field Head-related Transfer Functions.” *J. Acoust. Soc. Am*. Forthcoming.

Brungart, D.S., N.I. Durlach, and W.M. Rabinowitz. “Auditory Localization of Nearby Sources II: Localization of a Broadband Source.” *J. Acoust. Soc. Am.* Forthcoming.

Brungart, D.S., and N.I. Durlach. “Auditory Localization of Nearby Sources III: Stimulus Effects.” Submitted to *J. Acoust. Soc. Am*.

Shinn-Cunningham, B.G., N.I. Durlach et al. “Adapting to Supernormal Auditory Localization Cues I: Bias and Resolution.” *J. Acoust. Soc. Am.* 103(6): 3656-66 (1998).

Shinn-Cunningham, B.G., N.I. Durlach et al. “Adapting to Supernormal Auditory Localization Cues II: Changes in Mean Response.” *J. Acoust. Soc. Am.* 103(6): 3667-76 (1998).

Shinn-Cunningham, B.G. “Adapting to Remapped Auditory Localization Cues: a Decision-theory Model.” Submitted to *Percept. Psychophys*.

Thesis

Kassem, S. *Adaptation to Auditory Localization Cues from an Enlarged Head.* M.Eng. thesis, Depart-

40 B.G. Shinn-Cunningham, N.I. Durlach et al., “Adapting to Supernormal Auditory Localization Cues I: Bias and Resolution,” *J. Acoust. Soc. Am.* 103(6): 3656-66 (1998); B.G Shinn-Cunningham, N.I. Durlach et al., “Adapting to Supernormal Auditory Localization Cues II: Changes in Mean Response,” *J. Acoust. Soc. Am.* 103(6): 3667-76 (1998); B.G. Shinn-Cunningham, “Adapting to Remapped Auditory Localization Cues: a Decision-theory Model,” Submitted to *Percept. Psychophys.*; D.S. Brungart and W.M. Rabinowitz, “Auditory Localization of Nearby Sources I: Near-field Head-related Transfer Functions,” *J. Acoust. Soc. Am.*, forthcoming; D.S. Brungart, N.I. Durlach, and W.M. Rabinowitz, “Auditory Localization of Nearby Sources II: Localization of a Broadband Source,” *J. Acoust. Soc. Am.*, forthcoming; D.S. Brungart and N.I. Durlach, “Auditory Localization of Nearby Sources III: Stimulus Effects,” Submitted to *J. Acoust. Soc. Am.*

41 S. Kassem, *Adaptation to Auditory Localization Cues from an Enlarged Head*, M.S. thesis, Department of Electrical Engineering and Computer Science, MIT, 1998.

ment of Electrical Engineering and Computer Science, MIT, 1998.

1.10 Role of Skin Biomechanics in Mechanoreceptor Response

Sponsor

National Institutes of Health
Grant RO1-NS33778

Project Staff

Dr. Mandayam A. Srinivasan, Balasundar I. Raju, Suvranu De, Jhung-Chi Liao, Joshua P. Cysyk, Professor Robert H. LaMotte[42]

1.10.1 Overview

Mechanics of the skin and subcutaneous tissues is as central to the sense of touch as optics of the eye is to vision and acoustics of the ear is to hearing. 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. In spite of the fundamental importance of the sense of touch in our lives, very little is known about the mechanics and the mechanisms of touch.

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.

The research work consisted of four parts: (1) to develop two and three-dimensional (3-D) mechanistic models of the primate fingertip, and gradually refine them so that their geometrical and material properties are increasingly realistic; (2) to expand the variety of stimuli that are pressed or stroked on the models in simulations of neurophysiological experiments; (3) to perform a series of biomechanical experiments under *in vivo* conditions using a variety of techniques including the use of videomicroscopy, magnetic resonance imaging (MRI), high-frequency ultrasound, and computer controlled stimulators; (4) to collaborate with Professor LaMotte of Yale University School of Medicine in obtaining and analyzing peripheral neural response data for a variety of tactile stimuli. During the past year, we have built a novel device, the ultrasound backscatter microscope (UBM), which is capable of imaging the papillary ridges as well as skin layers underneath at much higher resolution than MRI. We propose to develop this unique device further and use it to both directly observe the mechanistic phenomena in tactile sensing as well as to measure the biomechanical parameters to help improve the realism of our models.

The progress described in the following sections are organized according to the research area: (1) biomechanics, (2) neurophysiology and psychophysics, (3) computational models, (4) theory, and (5) device design and construction.

1.10.2 Biomechanics

Determination of Compressibility and Mechanical Impedance of the Human Fingerpad *in Vivo*

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. Our results show that the compressibility of the fingertip increases with increases in both the depth of indentation and the contact area with the indentor. The highest change in fingertip volume was about 5%. We have also developed an experimental setup involving a computer controlled linear actuator for fingertip volume change measurements under dynamic

42 Yale University, New Haven, Connecticut.

conditions. 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 indentor displacement and are independent of velocity; during sawtooth stimulations, however, the nature of the hysteresis loops depend on velocity of indentation.

We have also measured the force response of the human fingerpad *in vivo* to indentation by stimuli of varying geometry. A computer-controlled tactile stimulator was constructed to deliver a combination of static, ramp and sinusoidal indentations normal the skin surface, with the fingerpads of subjects held stationary and passive. Both input indentation depth and fingerpad force response were recorded as functions of time to capture transients and steady state features. Three rigid metal indentors, a point, a 6.35 mm diameter circular probe and a flat plate, were used for indentation to represent three general classes of loading profiles encountered in manual exploration and manipulation. With each stimulus, repeatability of the response was tested and the effects of varying amplitude, velocity, and frequency of indentation were investigated.

The experiments revealed that the force response of the fingerpad is both nonlinear and viscoelastic with respect to indentation depth and velocity. A clear variation was present in the force response among the five subjects tested and across the three different indentors. This variation was minimized partly by determining a subject specific parameter with which to normalize the force response data. A nonlinear Kelvin model was then proposed to mathematically represent this characteristic force response of the fingerpad to indentation as a function of indentor and subject. However, in implementation, the nonlinear model was approximated by a lumped parameter model composed of a piecewise linear set of springs in parallel with series spring-dashpots. Parameters were estimated for the model for each subject and indentor given the experimental input and normalized output data. These “individual” models were able to predict data for that particular subject and indentor very well ($R^2 > 0.96$) but not as well for others. The means of the parameters across subjects were then utilized to construct more general, indentor specific versions of the model, which were able to predict better the force response of any subject’s fingerpad to a given indentation. These results were used in validating two-dimensional and three-dimensional mechanistic models of the primate fingertip.

Experimental Investigation of Frictional Properties of the Human Fingerpad

In manual exploration as well as manipulation, the frictional properties of the fingerpad play a dominant role in governing the forces applied, the amount of skin stretch, and the occurrence of slip. However, no data on the frictional characteristics of the primate fingerpad was available. Therefore we used the same tactile stimulator to indent and stroke the fingerpads of human subjects. It was programmed to deliver different indentation depths, stroke velocities, and stroke directions. Three flat plates made of glass, polycarbonate, and acrylic were used as stimulus surfaces in this experiment. During stroking, the normal and shear forces were recorded by a 2-axis force sensor. A videomicroscopy system was set up to capture the image sequences of the contact region between the fingerpad and the stimulus surface while stroking. The stimulator and the videomicroscopy system were synchronized so as to match the images with the corresponding force data. Five subjects participated in this experiment.

The data show distinct frictional behaviors for different stimulus surfaces. For the glass surface, the curves of normal as well as shear forces increased smoothly to steady state values. When the indentation depth was larger, the normal and shear forces were larger, but the friction coefficient was smaller. When the stroke velocity increased, the normal force was about the same for a given indentation depth, while the shear force and the friction coefficient increased. The stroke direction did not significantly influence the results. The image sequence shows that the relative motion, or slip, between the fingerpad and the glass plate indentor began at the periphery and propagated towards the center. The displacements of different finger ridges in the contact area were also different.

Polycarbonate and acrylic surfaces, although similar in smoothness and appearance to glass, caused a radically different frictional behavior: stick-slip phenomenon occurred consistently all through the stroke in every trial. An analysis of the stick-slip frequency and the stick-slip shear force was conducted with respect to various indentation depths and various stroke velocities. A hypothesis about junction forming rate and junction breaking rate was proposed based on adhesion theory. This was used to explain the different results from the glass plate and the polycarbonate plate. A comparison of the frictional properties of the fingerpad and rubber-like materials

shows the same trend in the data for the two cases. The frictional data we have obtained will be incorporated into our 3-D model of the primate fingertip to make the simulations of stroking of stimulus objects more realistic.

Investigation of the Internal Geometry and Mechanics of the Human Fingertip *In Vivo* using Magnetic Resonance Imaging

To gain insight into the mechanistic bases of the human tactile sensory system, we have developed a unique series of increasingly complex and detailed biomechanical models of monkey and human fingertips. These models are necessary to generate testable hypotheses on tactile neural coding or to predict neural response of individual or a population of mechanoreceptors. Although three-dimensional models of human and monkey fingertips with realistic external geometry and multilayered interior have been completed, the geometry and material properties of the internal layers have been idealized. Empirical data on the deformation behavior of the internal layers is essential for validating these models.

We employed advanced techniques in magnetic resonance imaging (MRI) to obtain realistic internal geometry and deformation of the tissue layers of the *in vivo* human fingerpad. The fingerpads of four subjects were statically loaded with various indentors to examine the effects of indentation depth and indentor shape on tissue deformation. Geometric surfaces that simulate edges and surfaces of objects, such as a line load, various sized rectangular bars and cylinders, were used to load the fingertip. Utilizing a 4.7 Tesla magnet and a RARE sequence, we were able to obtain images with 125 μm x 125 μm in-plane resolution. It should be noted that this resolution is very high compared to the typical MRI data obtained in clinical applications. Digital image processing techniques were used to filter the images and to detect the boundaries of the tissues located in the fingertip. Edge detection algorithms based on conformable contours ("snakes") allowed for separation of tissue layers. Published data on histology, and anatomy were used to identify each tissue layer in the fingertip.

The geometric information extracted from each tissue layer was used to examine tissue deformation during loading and is being used to improve the realism of the computational models. This data confirmed our earlier conclusions based on simulations that the soft tissues of the fingerpad act as low pass filters. The implication is that the high-spatial frequencies present in edges and corners are attenuated at the mechanoreceptor locations inside the fingerpad. Additionally, the results indicate that the fingerpad is compressible under load. Thus MRI proved to be a powerful tool to visualize soft tissue structures inside the fingerpad and their deformation behavior under *in vivo* conditions. The high resolution and contrast we were able to achieve enabled discrimination of tissue layers. It is therefore a useful tool for further investigation of the effects of static loading and, in the future, dynamic loading on the fingertip as a whole. However, a complementary imaging technique that has a higher resolution is needed to examine the details of the mechanical behavior of skin layers enclosing the mechanoreceptors. Therefore we built the ultrasound backscatter microscope.

1.10.3 Neurophysiology and Psychophysics

Tactile Coding of Shape

A salient feature of tactile sensing is its ability to encode and decode the shape of objects. In collaboration with Prof. LaMotte of Yale University School of Medicine, we have recorded the responses of SAIs and RAs to a variety of 2-D and 3-D shapes stroked across the monkey fingerpad. One set of experiments involved 2-D "wavy surfaces," i.e, surfaces composed of smooth, alternating convex and concave surfaces of differing radii of curvature. The second set of experiments employed 3-D toroidal objects mounted on a flat plate. With wavy surfaces, it was shown that only convexities were encoded in the neural responses; concavities evoked no responses. The primary findings from both sets of experiments were as follows: (1) discharge rates encode the magnitude and rate of change in the curvature of the skin produced by an object, (2) the orientation and shape of the two-dimensional outline of the object parallel to the skin are represented by the orientation and shape of the region of neural activity in both SA and RA populations, (3) object shape perpendicular to the skin is encoded in the shape of the object SA SPR (Spatial Population Response), (4) When object curvature is constant (e.g., circular cylinders), the slopes of the rising and falling phases of the SA response profile are constant, and (5) spatial measures of shape (width and average slope from base to peak) were generally found to be invariant with changes in the orientation of the object as well as the velocity and direction of stroking.

We have also investigated how shapes are encoded by populations of RAs and SAIs using a novel paradigm. 3-D toroidal objects were indented at a fixed location on the monkey fingerpad, and an estimate of the responses from a spatially distributed population of mechanoreceptors was obtained by successively recording single fiber responses and plotting the collection of responses on a "virtual" monkey fingerpad. This was a shift from the usual experimental paradigm where "population response" is estimated by applying the stimulus to various locations in the receptive field of a single afferent fiber. A major conclusion from these studies was that the stimulus shape and orientation were unambiguously coded by the Spatial Population Response Profiles (SPR) of SAs, while the RA SPR did neither. This shape code is expected to be essentially invariant with changes in force or velocity of indentation, as demonstrated for raised toroidal objects on a planar surface described above.

Tactile Coding of Softness

Encoding of softness is perhaps even more important in tactile sensing than that of shape, because softness can only be sensed accurately by direct touch whereas shape can be inferred through vision as well. We have described, for the first time, how primates discriminate between objects of different compliances and what is the biomechanical and neural basis of the perception of softness. We have shown that compliant springs with rigid surfaces ("spring-cells") required both kinesthetic and tactile information for softness discrimination, whereas for soft rubber objects of different compliances, tactile information alone was sufficient. This is because for a given force applied by a compliant object to the skin, the spatial pressure distribution and skin deformation within the contact region depend on the specimen compliance if the object has a deformable surface (e.g., fruits), but is independent of the specimen compliance if its surface is rigid (e.g., piano key). Thus, tactile information alone is necessary and sufficient to encode the compliance of rubber-like objects.

We then focused on finding a more quantitative neurophysiological and biomechanical basis for softness encoding. Using a computer-controlled tactile stimulator, we applied rubber specimens to the fingerpads of anesthetized monkeys in a controlled manner and recorded the neural response from SAI and RA fibers. The discharge rates were observed to be lower in the SAI fiber's response to softer specimens compared to the more stiff ones. In contrast, RA response was found to be practically indifferent to the relative variations in stiffness. Thus, it was concluded that tactile discrimination of softness was based more on the discharge rates from the SAIs than from the RAs. It was also found that when specimens were applied to the fingerpad at the same velocity, the softer the specimen, the lower the rate of change of net force and the higher the rate of change of overall contact area. Thus at a given instant of time during indentation, the difference in the average pressure between the two specimens was higher than the corresponding differences in either the forces or the contact areas. Just as the pressure increased more slowly for the softer specimen, the SA discharge rate also increased more slowly, resulting in a slower increase in cumulative impulses. However, the force, contact area, and discharge rate were affected by the velocity of indentation. For the same specimen, the lower indentation velocity resulted in lower force and area rates, giving rise to a lower discharge rate at a given instant of time during the ramp. Since the discharge rate of a single fiber is affected by both the compliance of the specimen and the indentation velocity, specimens of differing compliances could be made to give rise to the same single fiber response by appropriate adjustment of indentation velocity. Thus, discharge rate in a single SAI fiber cannot unequivocally encode the compliance of an object, but a population of spatially distributed SAIs can.

1.10.4 Computational Models

In order to better understand the mechanics of touch, it is necessary to establish a quantitative relationship between the stress/strain state at a mechanoreceptor location and the neural response of the receptor to a given mechanical stimulus. Due to the subsurface locations of the receptors and the opacity of the skin, the stress state and deformations in the close vicinity of a receptor cannot be observed experimentally. Moreover, no experimental techniques exist to record the responses from a population of mechanoreceptors. A mechanistic model of the skin and subcutaneous tissues that is validated through biomechanical and neurophysiological experiments is able to establish the stress/strain stimulus to a mechanoreceptor as well as predict the population response to a given stimulus. Therefore, we developed a series of increasingly realistic 2-D and 3-D finite element models of the primate fingertip. We summarize below the development of the 3-D model and the biomechanical and neurophysiological results obtained from it.

Development of Three-Dimensional Layered Model of Human and Monkey Fingertips

The external geometry of human and monkey fingertips was obtained from precise epoxy casts made using dental cement molds. These casts were extremely accurate in reproducing the finger print ridges, details of the nail and wrinkles on the skin. A videomicroscopy setup consisting of a monochrome CCD camera with zoom lenses, a frame grabber, and a PC was used to acquire images of the casts in different orientations. A stepper motor was used to rotate the fingertip about an axis parallel to the bone axis in 1 degree steps, and an image was grabbed at each step. The boundary of the fingertip in an image frame essentially represented the orthographic projection of the fingertip for that particular orientation. These 2-D sections were imported into a solid modeler software (PATRAN) and a 3-D model of the fingertip with realistic external geometry was generated. The relative thickness of the bone in the distal phalanx was determined from X-ray images and a concentric bone was generated inside the fingertip. To account for the several layers of skin and the adipose tissue underneath, the mesh was generated in layers such that each layer could be assigned a distinct material property and mechanistic constitutive behavior. The material of each layer was treated as linear isotropic and the innermost layer was made several orders of magnitude stiffer than all the other layers to simulate the rigid behavior of the bone. Two models with eight-noded isoparametric elements were generated and the number of nodes in the two models were 8500 and 30,000 respectively. The typical diameter of the monkey fingertips was approximately 9 mm and element size in the region of contact with indentors was approximately 500 microns and 160 microns for the two models respectively.

Encoding and Decoding of Shape during Static Tactile Sensing

The fingertip model described above was used to simulate static indentation of the fingertip by rigid objects of different shapes such as cylinders rectangular bars, and sinusoidal step shapes. The large number of numerical computations necessary to achieve a high spatial resolution and realism in the simulations required the use of a supercomputer (Cray C90). The results show that contact mechanics is important in governing the pressure distribution on the skin surface, which, in fact, is the stimulus unique to each shape. This surface pressure distribution within contact regions was found to be highly dependent on the curvature of the object that indented the finger. Further, we have shown that a simple equation is able to predict the surface pressure as a function of the indenting object's curvature and the local depth of indentation. To study the mechanism of transduction by the mechanoreceptors (transformation of the mechanical stress state into neural signals), 21 mechanical measures were obtained from the calculated stress and strain tensor at mechanoreceptor locations, and were matched with experimentally recorded neural response data. Three quantities—maximum compressive strain, maximum tensile strain and strain energy density—were found to be related to the neural responses of SA-I nerve fibers through a simple scaling-threshold model and are thus possible *relevant stimuli* for SA-I afferents. Among these, strain energy density is more likely to be the relevant stimulus since it is a scalar that is invariant with respect to receptor orientations and is a direct measure of the distortions of the receptor caused by the loads imposed on the skin.

To identify the object contacting the skin, the CNS should be able to compute surface loads imposed on the skin from the peripheral neural response. To simulate this inverse problem of decoding, a nonlinear shift-invariant system, which treats the surface pressure as input and neural responses as output, was developed. Because of the nonlinearity (the relevant stimulus measures, such as the strain energy density, are nonlinear functions of the cartesian stress-strain components), a simple inverse transformation cannot be applied. A signal estimation technique using the univariate method used in non-linear optimization techniques was employed to decode the surface pressure function from the neural response function. The decoding was demonstrated to be valid for both the ideal case where no sensor noise is present as well as the case where the sensor noise (assumed to be additive Gaussian) is present, as long as the signal-to-noise ratio is greater than 20 dB. This result shows a method by which the central nervous system could infer the shape of the object contacting the skin from SAI population response under static conditions.

Modeling the Dynamics of the Primate Fingerpad

The previous section describes our fingertip models that are able to explain and predict both biomechanical and neurophysiological phenomenon observed in experiments with static stimuli. Encouraged by this

success, we have now begun to model the dynamic behavior of the fingerpad in order to realistically simulate the neurophysiological experiments involving dynamic stimuli, such as under stroking of shapes. We have now incorporated viscoelasticity into our computational models of the primate fingertip. To this end, the biomechanical data obtained from the indentation of the fingerpads of several human subjects using different indentor geometries was utilized. A consistent normalization scheme was developed which showed that most of the variation in the data obtained across subjects was scalable by a single parameter. This lead to the development of a second order Kelvin model which satisfactorily explains much of the observed force-displacement data for a truncated conical indentor. The correspondence principle was invoked to extend these results to obtain the material parameters of a generalized 3-D linear viscoelastic continuum. These parameters were then incorporated into a 2-D plane strain and a 3-D layered finite element model. The results obtained from these computational models predict the observed force-displacement data very well for all the indentors (truncated conical, cylindrical and flat-plate indentors) used in the earlier biomechanical experiments. These models are now being used to simulate dynamic stimuli imposed on the fingerpad, such as stroking of shapes in order to understand the role of mechanoreceptors during haptic exploration. Neurophysiological recordings from slowly adapting (SA) and rapidly adapting (RA) mechanoreceptors have been made for a variety of shapes, both statically indented and dynamically stroked across the fingerpad. Previous biomechanics research has been to determine the mechanics underlying the role of SAs during static indentation. Mechanical cues have been determined which relate curvature to impulse response of the receptor. The purpose of the current investigation is to determine the mechanical response of both SAs and RAs during dynamic stroking, and to develop a unifying model of the role of each mechanoreceptor in touch sensation.

Using MRI images of human fingertips, a two-dimensional model of the fingerpad was developed for finite element analysis. This model accurately depicts the inner structure of the finger. The material properties of each layer of the fingerpad were obtained from previous biomechanical studies. The MRI images provided detailed knowledge of the inner layer deformation profiles during static indentation. Static indentations on the model were simulated, and the deformation profiles of the inner and outer layers were compared to MRI images in order to validate the model.

Once the model was validated, stroking was simulated using surfaces of various curvatures, stroked at different velocities. The biomechanical properties of the fingerpad were investigated using these simulations; the relationship between contact pressure and mechanical response at the mechanoreceptor depth was examined. Research is in progress to determine the relevant mechanical stimulus for both RA and SA type receptors, and ultimately relate the current mechanical state of the fingerpad to the neurophysiological response of the mechanoreceptors.

1.10.5 Theory

Nonlinear Dynamics of Mechanoreceptor Response

One of the most interesting aspects of dynamic tactile sensing in humans is the nature of mechanoreceptor response to dynamic stimuli. In contrast to the response of the fingerpad tissue, the receptors seem to exhibit nonlinear behavior even for very small indentations of the fingerpad skin.

The most classic example of such nonlinear response is the so called “tuning curves” which are nothing but the variations of dead-zone and saturation thresholds as functions of frequency of input sinusoids. In order to model these nonlinearities, a generalized class of cascaded LNL-type filter banks were developed. Such models, in general, incorporate a linear describing function block followed by a static nonlinearity and another linear describing function block. It was observed that different receptor classes could be described by specializing this general model. For instance, the behavior of the SAI mechanoreceptors could be explained very well using a Hammerstein type of structure (a static nonlinearity followed by a linear dynamic block). These models provided good fits to the empirically recorded mechanoreceptor responses. The next step appears to be a successful link between the finite element model describing the geometric and material properties of the fingerpad and the neuro-dynamic transduction blocks, describing receptor behavior for each class of receptors. 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) on primate fingerpads.

Identification and Control of Haptic Systems: A Computational Theory

This research provides a theoretical framework for haptics, the study of exploration and manipulation using hands. In both human and 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. We first investigated 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.

We have shown that three quantities, mean normal stress and the two shear strains at mechanosensor locations, are not only necessary and sufficient to infer the surface signals, but also maximize the spatial bandwidth of signal reconstruction. We then focused on how the information obtained from such optimal sensing can be used for exploration of objects. We have shown that an accurate reconstruction of object properties can occur using two basic building blocks of Exploration Strategy and Finger Control. Exploration Strategy pertains to the problem of inferring object properties such as shape, texture and compliance, and interference properties such as state of contact, from the estimated surface signals. This involves determining, in each case, what kind of sensor information and what kind of action is needed. Finger Control refers to the transformation of the action needed into a command trajectory for the fingerpad, which defines the desired direction of movement for manipulation. We have defined and analyzed the components of both these blocks, provided explicit mathematical formulation, and have solved 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.

1.10.6 Device Design and Construction

Ultrasound Backscatter Microscope for In Vivo Imaging of Human Fingertip

One of the conclusions of our earlier MRI studies was that if a noninvasive imaging system with higher resolutions than MRI could be designed, it would be a powerful tool to empirically observe the mechanical deformations of the skin tissue around mechanoreceptors and would help validate our computational models. We have now developed an ultrasound backscatter microscope (UBM), which is able to display the geometry and deformation of skin layers under *in vivo* conditions. UBM is similar to B-mode diagnostic ultrasound imaging, but uses higher frequency acoustic waves (about 50 MHz) to achieve resolutions of the order of tens of microns. In UBM, contrast depends on the mechanical properties of tissues, a feature that complements techniques such as optical microscopy, CT and MRI that rely on other tissue properties. This feature also makes UBM ideal for studying the mechanistic basis of tactile sensing. In addition, UBM is less expensive than most imaging techniques, and is also noninvasive. However, because of increased attenuation of the acoustic waves at higher frequencies, the tissues being imaged must be located within a few millimeters of the surface. A UBM system was designed and built using a high-frequency PVDF transducer (nominal frequency of 75 MHz), a pulser, a digitizing oscilloscope, a scanning system and the IEEE488 interface.

The device was used to image human fingertip skin under *in vivo* conditions so as to obtain information about the internal structure of the fingertip. At each skin location, the transducer was energized and echoes from tissues at different depths were recorded. By mechanically scanning the transducer across the fingerpad surface and keeping track of signals from successive lateral locations, data on mechanical contrast in skin cross sections were assembled. Signal processing was done on the reflected echoes to obtain 2-D images. Images of fingerpad skin of six human subjects showed three distinct layers up to a depth of about 1.2 mm. Comparison images of fin-

gertip skin on the dorsal side also showed a layered structure, with lesser thicknesses for the first two layers. The data obtained are consistent with known anatomical information in that the three layers imaged are the stratum corneum, the rest of the epidermis, and the top region of the dermis. These are the skin layers where the Meissner Corpuscle and the Merkel cells are known to be present. Although the current resolutions (150 microns laterally x 20 microns axially) of the UBM are sufficient for now, we believe these can be cut down to 30 microns x 10 microns with moderate changes. Even better resolutions may be possible with more complex calibration and signal processing. The UBM is designed to be portable, and therefore it may be possible to use it to image the deformation around a mechanoreceptor location while simultaneously recording its electrophysiological response.

During the past year: (1) we have made improvements in the hardware of the system and (2) we have developed theoretical models for tissue discrimination. The improvements in hardware consisted of the use of better-focused transducers, and the design of a high precision computer-controlled scanning system. The previous transducer employed had an f/# (ratio of focal length to diameter) of 4, which provided a lateral resolution of only 150 microns. Subsequently we have employed a more focused transducer with an f/# of 2 that yielded a lateral resolution of 80 microns. With better lateral resolution we were able to more clearly identify the finger ridges on the surface of the fingertip. During the year, we also designed a 3-axis computer-controlled scanning system that has a positioning accuracy of 1 micron. The system is currently being assembled. When completed it is expected that the time of imaging will be fast (few seconds for a 2-D scan) enough to minimize the motion artifacts that are present in the current manually scanned system.

Our efforts over the past year have also been to develop signal processing algorithms for tissue discrimination–differentiating one tissue from another. Unlike conventional image processing algorithms that rely only on the final image, our methods will utilize the properties of the tissues (scatterer size, scatterer number density), and the interaction of acoustic waves with tissues (speckle formation), which are not directly evident in the final image. We are currently studying these acoustic wave-tissue interaction processes using computer simulations and imaging of tissue-mimicking phantoms. It is expected that these studies will have potential application not only in fingertip imaging but also in the non-invasive detection of cutaneous melanoma.

High-Precision Tactile Stimulator for Dynamic Stimulation of the Fingerpad

Although our 2-axis tactile stimulator is still operational, its force and positional resolution as well as bandwidth were found to be in need of improvement in order to determine the effect of skin viscoelasticity on RA responses at stimulation frequencies higher than about 20 Hz. A high-precision tactile stimulator was built to allow us to refine the dynamic fingerpad model. The actuator is a disk-drive head-positioning motor coupled with angular position feedback from a precision rotary variable differential transformer (RVDT). A digital, real-time PID controller is implemented using a floating-point DSP system. Any indentor shape can be attached to the device to provide one-degree of freedom position-controlled inputs in either the normal or tangential direction to the fingerpad surface. The force response is measured with a piezoelectric transducer that deflects only 2 nm for the maximum expected force of 2 N. The stimulator operates smoothly in the bandwidth of 0.1 Hz to 300 Hz. Positions can be resolved to 600 nm while forces can be resolved to 3 mN. To determine the mechanical properties of the human fingerpad *in vivo*, two linearly independent experiments in the linear range of the tissue were performed to solve for the dynamic, viscoelestic mechanical properties of the human fingerpad. The impulse response bulk modulus and the impulse response shear modulus were computed as a function of frequency. The impulse response bulk modulus varied from 55 kPa at 2 Hz to 8 Mpa at 200 Hz. The impulse response shear modulus varied from 90 kPa at 2 Hz to 8 Mpa at 100 Hz.

Journal Articles

Khalsa, P.S., R.M. Friedman, M.A. Srinivasan, and R.H. LaMotte. “Encoding of Shape and Orientation by Populations of Slowly and Rapidly Adapting Mechanoreceptors: I. Objects Indented into the Finger Pad.” *J. Neurophys.* Forthcoming.

Karason, S, M.A. Srinivasan, and A.M. Annaswamy. “Encoding and Decoding of Static Information in Tactile Sensing Systems.” *Int. J. Robot. Res.* Forthcoming.

LaMotte, R.H., M.A. Srinivasan, C. Lu, P.S. Khalsa, and R.M. Friedman. “A Raised Object on a Planar Surface Stroked across the Finger Pad: I.

Responses of Cutaneous Mechanoreceptors to Shape and Orientation." *J. Neurophys.* Under revision.

Chapter in a Book

Annaswamy, A.M., and M.A. Srinivasan. "The Role of Compliant Finger Pads in Grasping and Manipulation: Identification and Control." In *Essays on Mathematical Robotics,* The IMA volumes in *Mathematics and its Applications*, vol. 104. J. Baillieul, S. Sastry, and H.J. Sussman, eds. New York: Springer-Verlag. Forthcoming.

Theses

Birch, A.S. *Experimental Determination of the Viscoelastic Properties of the Human Fingerpad*. S.M. thesis. Department of Mechanical Engineering, MIT, June 1998.

Raju, B.I. *Ultrasound Backscatter Microscope for In Vivo Imaging of Human Fingertip*. S.M. thesis. Department of Electrical Engineering and Computer Science, MIT, February 1998.

Raju, B.I. *Encoding and Decoding of Shape in Tactile Sensing*. S.M. thesis. Department of Mechanical Engineering, MIT, February 1998.

1.11 Force-Reflecting Soft Tissue Models for Simulating Laproscopic Surgical Procedures in Virtual Reality Systems

Sponsors

Harvard Medical School, Center for Minimally Invasive Surgery

Massachusetts General Hospital, Center for Innovative Minimally Invasive Therapy

Project Staff

Dr. Cagatay Basdogan, Dr. Mandayam A. Srinivasan, Chih-Hao Ho, Suvranu De, Dr. Stephen D. Small,[43] Dr. David Rattner[44]

Overview

Research in the area of computer assisted surgery and surgical simulation has mainly focused on developing 3-D geometrical models of the human body from 2-D medical images, visualization of internal structures for educational and preoperative surgical planning purposes, and graphical display of soft tissue behavior in real time. Conveying to the surgeon the touch and force sensations with the use of haptic interfaces has not been investigated in detail. We have developed a set of haptic rendering algorithms for simulating "surgical instrument-soft tissue" interactions. Although the focus of the study is the development of algorithms for simulation of laparoscopic procedures, the developed techniques are also useful in simulating other medical procedures involving touch and feel of soft tissues. The proposed force-reflecting soft tissue models are in various fidelities and have been developed to simulate the behavior of elastically deformable objects in virtual environments. The developed algorithms deal directly with geometry of anatomical organs, surface and compliance characteristics of tissues, and the estimation of appropriate reaction forces to convey to the user a feeling of touch and force sensations.[45]

The hardware components of the set up include a personal computer (300 MHz, dual Pentium processor) with a high-end 3-D graphics accelerator, a force-feedback device (PHANToM from SensAble Technologies Inc.) to simulate haptic sensations. During the simulations, the user manipulates the generic stylus of the force-feedback device to simulate the movements of a surgical instrument and to feel its interactions with the computer generated anatomical organs. The associated deformations of the organs are displayed on the computer monitor and reaction forces are fed back to the user through the haptic interface. The software was written in C/C++, using multithreading techniques to create separate visual and haptic control loops, thereby increasing the haptics servo rate (varies from 500 Hz to 2 kHz) while simultaneously satisfying the requirements of graphics update rate of at least 30 Hz.

43 Massachusetts General Hospital, Boston, Massachusetts.

44 Harvard Medical School, Boston, Massachusetts.

45 C. Basdogan, C. Ho, M.A. Srinivasan, S.D. Small, and S.L. Dawson, "Force Interactions in Laparoscopic Simulations: Haptic Rendering of Soft Tissues," Paper published in the *Proceedings of the Medicine Meets Virtual Reality (MMVR '98) VI Conference,* San Diego, California, January 19-22, 1999, pp. 385-91.

1.11.1 Progress During the Past Year

During the past year, we have made significant progress in the following areas:

(1) Development of "thin-walled" models to simulate tissue deformations and to compute reaction forces: In the language of mechanics, the "thin-walled" structures are broadly classified into membranes, structures with essentially no bending stiffness compared to the in-plane stiffness, and shells, structures in which bending behavior is also important. Triangular elements are used to represent the organ geometry and the virtual work principle is used to derive the incremental equations of motion.[46] The initial results suggest that the suggested "thin-walled" models can predict nonlinear behavior of tissues.

(2) Development of algorithms to simulate tissue cutting and bleeding: Realistic simulation of tissue cutting and bleeding are important components of a surgical simulator that are addressed in this study. Surgeons use a number of instruments to perform incision and dissection of tissues during minimally invasive surgery. For example, a coagulating hook is used to tear and spread the tissue that surrounds organs and scissors are used to dissect the cystic duct during laparoscopic cholecystectomy.

During the execution of these procedures, bleeding may occur and blood flows over the tissue surfaces. We have developed computationally fast algorithms to display (1) tissue cutting and (2) bleeding in virtual environments with applications to laparoscopic surgery. Cutting through soft tissue generates an infinitesimally thin slit until the sides of the surface are separated from each other. Simulation of an incision through tissue surface is modeled in three steps: first, the collisions between the instrument and the tissue surface are detected as the simulated cutting tool passes through. Then, the vertices along the cutting path are duplicated. Finally, a simple elastic tissue model is used to separate the vertices from each other to reveal the cut.

Accurate simulation of bleeding is a challenging problem because of the complexities of the circulatory system and the physics of viscous fluid flow. There are several fluid flow models described in the literature, but most of them are computationally slow and do not specifically address the problem of blood flowing over soft tissues. We have reviewed the existing models, and have adapted them to our specific task. The key characteristics of our blood flow model are a visually realistic display and real-time computational performance. To display bleeding in virtual environments, we developed a surface flow algorithm. This method is based on a simplified form of the Navier-Stokes equations governing viscous fluid flow. The simplification of these partial differential equations results in a wave equation that can be solved efficiently, in real-time, with finite difference techniques. The solution describes the flow of blood over the polyhedral surfaces representing the anatomical structures and is displayed as a continuous polyhedral surface drawn over the anatomy.[47]

Meeting Papers

Basdogan, C., C. Ho, and M.A. Srinivasan. "Simulation of Tissue Cutting and Bleeding for Laparoscopic Surgery Using Auxiliary Surfaces." Paper published in the *Proceedings of the Medicine Meets Virtual Reality Conference*, San Diego, California, January 19-22, 1999.

De, S., and M.A. Srinivasan. "Thin-Walled Models for Graphical and Haptic Rendering of Deformable Objects." Paper published in the *Proceeding of the Third PHANToM Users Group Workshop*, Dedham, Massachusetts, October 3-4, 1998.

De, S., and M.A. Srinivasan. "Thin-Walled Models for Haptic and Graphical Rendering of Soft Tissues in Surgical Simulations." *Proceedings of Medicine Meets Virtual Reality Conference 7*, San Francisco, California, January 1999.

46 S. De and M.A. Srinivasan, "Thin-Walled Models for Graphical and Haptic Rendering of Deformable Objects," paper published in the *Proceeding of the Third PHANToM Users Group Workshop*, Dedham, Massachusetts, October 3-4, 1998.

47 C. Basdogan, C. Ho, M.A. Srinivasan, S.D. Small, and S.L. Dawson, "Force Interactions in Laparoscopic Simulations: Haptic Rendering of Soft Tissues," Paper published in the *Proceedings of the Medicine Meets Virtual Reality (MMVR '98) VI Conference,* San Diego, California, January 19-22, 1999, pp. 385-91.