Section 2 Sensory Communication

Chapter 1 Sensory Communication

Chapter 1. Sensory Communication

Academic and Research Staff

Professor Louis D. Braida, Nathaniel I. Durlach, Professor Richard M. Held, Dr. Paul Duchnowski, Dr. Julie E. Greenberg, Dr. Karen L. Payton, Dr. Matthew H. Power, Dr. William M. Rabinowitz, Dr. Christine M. Rankovic, Dr. Charlotte M. Reed, Dr. J. Kenneth Salisbury, Dr. Barbara G. Shinn-Cunningham, Dr. Mandayam A. Srinivasan, Dr. Thomas E.v. Wiegand, Dr. Patrick M. Zurek, Merry A. Brantley, Andrew R. Brughera, Lorraine A. Delhorne, Joseph G. Desloge, Joseph A. Frisbie, Dorrie Hall, Seth M. Hall, Michael P. O'Connell

Visiting Scientists and Research Affiliates

Dr. Cagatay Basdogan, Dr. Susan L. Goldman, Dr. Kenneth W. Grant, Dr. Kourosh Saberi, Dr. Kaoru Sekiyama, Dr. Annie H. Takeuchi, Geoffrey L. Plant

Graduate Students

Amanda S. Birch, Andrew G. Brooks, Douglas S. Brungart, Joshua P. Cysyk, Suvranu De, Matthew E. Esch, Isaac J. Graf, Chi-Hao Ho, Alexandra I. Hou, Salim F. Kassem, Glenn Koh, Jean C. Krause, David S. Lum, Sharieff A. Mansour, Lajos Molnar, Sadiki P. Mwanyoha, Balasundar I. Raju, David W. Schloerb, Steven J. Schlueter, Matthew G. Sexton, Jason J. Sroka, Maciej Stachowiak, Prasanna B. Tambe, Xudong Tang, Dinh Yen Tran, Kimberly J. Voss, Evan F. Wies, Wan-Chen Wu, Hanfeng Yuan, Jonathan L. Zalesky

Undergraduate Students

David E. DiFranco, Brian D. Kincy, Kari Anne H. Kjolaas, Samuel R. Madden, David H. Manowitz, Adrienne H. Slaughter

Technical and Support Staff

Rebecca L. Garnett, Eleanora M. Luongo

1.1 Introduction

The Sensory Communication group is interested in understanding sensorimotor and cognitive processes and applying this understanding to solve practical problems in a variety of domains. Our basic research is characterized by behavioral (psychophysical) experimentation and quantitative theoretical modeling. Facilities to support this research are developed as needed.

Although some research is conducted on vision, most of the group's work is focused on audition and taction. The main applications are concerned with aiding individuals who suffer from impaired hearing, developing improved human-machine interfaces for virtual environments and teleoperation (virtual reality), and using virtual environment technology for training. The main facilities of the group are associated with the psychoacoustics, touch (or haptics), and virtual environment laboratories.

1.2 Hearing Aid Research

Sponsor

National Institutes of Health
Grant R01 DC00117

Project Staff

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

1.2.1 Specific Aims

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

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

1.2.2 Studies and Results

Characteristics of Sensorineural Hearing Impairment

Simulation of Sensorineural Hearing Loss

In order to obtain insight into the perceptual effects of hearing loss, we have been developing *functional* models of hearing impairment. Such models would allow listeners with normal hearing to experience the perceptual distortions and speech-reception problems experienced by hearing-impaired listeners. Current research is focused on simulating the effects of abnormal loudness growth and reduced frequency selectivity.

Graf[1] processed sounds using combinations of multiband amplitude expansion (E) and spectral smearing (SS) to simulate the effects of sensorineural hearing impairment on tone-detection thresholds, dependence of loudness level on intensity, and frequency selectivity. The E simulation[2] was adjusted to produce a flat 40 dB loss. The SS simulation was adjusted to produce excitation patterns in normal ears similar to those produced by unprocessed stimuli in impaired ears with auditory filters whose bandwidths are increased by factors of 1.7 and 3.0, either symmetrically about the center frequency or asymmetrically to simulate auditory filters with reduced low-frequency slopes. In the psychoacoustic tests, smearing was achieved by replacing tonal stimuli with narrow bands of noise and by broadening the spectra of noise stimuli. In tests with speech stimuli, smearing was achieved by convolving the short-term mel-frequency power spectrum with a low-pass filter. The effects of the E and SS simulations in isolation and in combination were investigated in both psychoacoustic and speech reception tasks.

Both E and SS have only minor effects on in-band masking produced by narrowband noise. E produces only modest amounts of out-of-band masking, primarily for frequencies below the noise. SS increases out-of-band masking, with the symmetry and amount of masking reflecting the altered shape of the auditory filter associated with the smearing operation. E elevates the tip and decreases the slopes of the tails of psychoacoustic tuning curves; SS has little effect on the tip, but decreases the slope of one (asymmetric) or both (symmetric smearing) of the tails of tuning curves. Combinations of E and SS can approximate a range of narrowband masking patterns, such as those that exhibit near-normal low-frequency tails, but greatly increased upward spread of masking. E substantially reduces loudness summation, but SS has only minor effects on loudness summation, either alone or in combination with E.

The effects of the simulations on speech reception were evaluated by measuring the intelligibility of CV syllables both in quiet and in the presence of speech-shaped noise (e.g., Zurek and Delhorne[3]). In quiet, E reduces scores more than SS, with the reductions nearly adding when the simulations are combined. In noise, SS reduces scores (relative to the unsimulated condition) increasingly as the speech to noise ratio (SNR) is reduced, but E has little effect. When combined with SS, E had little effect on scores at +6 dB SNR, but, surprisingly, improved scores at 0 dB SNR.

Graf also explored whether SS could improve simulations of speech reception that were not well simulated by E alone. Duchnowski and Zurek,[4] who used E to simulate the impairment of one such listener

1 I.J. Graf, *Simulation of the Effects of Sensorineural Hearing Loss*, M.Eng. thesis, Department of Electrical Engineering and Computer Science, September 1997.

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

3 P.M. Zurek and L.A. Delhorne, "Consonant Reception in Noise by Listeners with Mild and Moderate Hearing Impairment," *J. Acoust. Soc. Am.* 82: 1548-59 (1987).

(AL) with a severe (65-80 dB) flat loss, found that consonant recognition was well simulated in both quiet and noise when a flat frequency-gain characteristic was used. But when high-frequency emphasis was used, the simulated normal listeners achieved higher intelligibility scores than AL. To improve the simulation Graf combined E with SS using E parameters identical to those used in the Duchnowski and Zurek[5] study and SS parameters that are generally representative of listeners with severe losses (AL being unavailable for further testing, auditory filter bandwidths were increased by a factor of three as suggested by Nejime and Moore.[6] The combined simulation eliminated the relative benefit provided by high-frequency emphasis to the listeners with simulated losses, but in all conditions it produced consonant intelligibility scores that were lower than those achieved by AL, suggesting that the amount of spectral smearing was too large.

Characteristics of the Speech Signal

Clear Speech

In adverse listening conditions, talkers can increase their intelligibility by speaking clearly. While producing clear speech, however, talkers often significantly reduce their speaking rate. A recent study[7] showed that talkers can be trained to produce a form of clear speech at normal rates (clear/normal speech). This finding suggests that acoustical factors other than reduced speaking rate are responsible for the high intelligibility of clear speech. To gain insight into these factors, the acoustical properties of conversational, clear/slow, and clear/normal speech are being analyzed to determine phonological and phonetic differences between the two speaking modes.

Acoustic Characteristics

Since a form of clear speech can exist at normal speaking rates, acoustic properties other than speaking rate must account for the high intelligibility of clear speech. We are currently attempting to identify these acoustic factors using two strategies: conducting additional intelligibility tests and measuring the acoustic properties of clear and conversational speech. Initial acoustic measurements have consisted of pause length and distribution, long-term spectra, short-term spectra, average duration and power of phones, and number of occurrences of phonological phenomena such as vowel modification, burst elimination, alveolar flaps, and sound insertions. These measurements of conversational and clear/slow speech have revealed differences that are consistent with those of Picheny et al.[8] and Uchanski et al.,[9] while initial measurements of conversational and clear/normal speech have revealed that the spectra of clear/normal speech contains more energy above 1 kHz than conversational speech.[10]

Additional intelligibility tests have also been developed for other degraded conditions in order to investigate which acoustic characteristics of the clear speech signal are responsible for its high intelligibility. These tests can also help evaluate the robustness of the high intelligibility of clear/normal speech. Three degradations have been explored with normal-hearing listeners: reverberation, low-pass filtering, and high-pass filtering conditions. In addition, non-native listeners were tested in additive noise. Of the two talkers tested, the results showed that one talker's clear/normal speech was more intelligible than his conversational speech for three of the conditions (non-native listeners: 14%; reverberation: 19%; high-pass: 19%). The only condition in which his clear/normal speech did not provide an intelligibility advantage was for the low-pass filtering condition. In

4 P. Duchnowski and P.M. Zurek, "Villchur Revisited: Another Look at AGC Simulation of Recruiting Hearing Loss," *J. Acoust. Soc. Am.* 98: 3170-81 (1995).

5 Ibid.

6 Y. Nejime and B.C.J. Moore, "Simulation of the Effect of Threshold Elevation and Loudness Recruitment Combined with Reduced Frequency Selectivity on the Intelligibility of Speech in Noise," *J. Acoust. Soc. Am.* 102(1): 603-15 (1997).

7 J.C. Krause, *The Effects of Speaking Rate and Speaking Mode on Intelligibility,* M.S. thesis, Department of Electrical Engineering and Computer Science, MIT, September 1995.

8 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 (1986).

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

10 J.C. Krause and L.D. Braida, "Properties of Naturally Produced Clear Speech at Normal Speaking Rates," *J. Acoust. Soc. Am.* 100: S2828 (1996).

this condition, his conversational speech was so highly intelligible (87%) that a ceiling effect could account for his failure to improve his intelligibility. The second talker's clear/normal speech, however, was significantly more intelligible than her conversational speech for only one of the four conditions tested (low-pass: 15%) tested.[11] The first talker's clear/normal speech appears to be more robust to other degradations than the second talker's clear/normal speech, suggesting that the two talkers have different strategies for producing clear/normal speech. Further acoustical measurements such as fundamental frequency distribution, consonant-vowel ratio, and vowel formant frequencies must be taken in order to investigate differences in the talkers' speaking strategies.

Computational Model of Speech Intelligibility based on the STI

In order to estimate the effects of a variety of types of signal processing transformations on speech, we are developing methods of predicting speech intelligibility from physical measurements made on speech waveforms. One method, the speech-based STI model, assumes that changes in envelope modulation spectra result in intelligibility changes.

The STI[12] is usually computed on the basis of acoustic theory or from measurements of changes to the envelope spectra of non-speech signals, such as amplitude-modulated noise bands. This limits the applicability of the STI when it is of interest to predict the effects of nonlinear hearing aid processing strategies such as amplitude compression, speaking style, and hearing impairment. To extend the STI to such situations, it is necessary to use speech waveforms as probes and to determine the relation between envelope spectra measured both before and after the speech is altered.

Direct determination of changes to the envelope spectra of bands of speech has proven sensitive to the effects of artifacts that arise when speech is degraded by noise and/or reverberation. Although the traditional theory-based STI computation makes fairly accurate predictions of the effects of such degradations on intelligibility, it has proven difficult to determine the STI with comparable accuracy from direct measurements using speech waveforms. The modulation spectra of degraded speech often exhibit artifacts that lead to modulation transfer functions (MTFs) that are inconsistent with acoustic theory. In particular, the MTFs for speech degraded by additive noise are often not attenuated uniformly as expected, most likely because the noise contributes its own modulations to the summed waveforms. Similarly, the MTFs for speech degraded by reverberation often do not exhibit the expected lowpass characteristic, most likely as the result of constructive and destructive interference between the direct and reflected speech waveforms. This interference varies as the relation between the arrival times of reflections and the fundamental period of the voice changes, producing unexpected fluctuations in the speech envelope.

To reduce the effects of these artifacts on the computation of the STI, determine the coherence between the envelopes of the degraded and undegraded speech waveforms. The coherence is typically close to unity when artifacts are absent, but drops towards zero when artifacts are in evidence. Inspection of coherence functions for speech that was subject to a variety of degradations suggested that only modulations below the critical frequency at which the coherence fell below 0.8 were likely to be free of the effects of artifacts. This led us to specify an upper bound on the envelope modulation frequencies that are included in the computation of the modulation transfer function from speech waveforms.[13] Our estimates of the STI derived from speech waveforms are monotonically related to intelligibility scores for the same speech waveforms when degraded by noise, reverberation, and combinations of noise and reverberation.[14] This result validates our procedures for computing the STI using speech waveforms as probes, at least for degradations consisting of additive noise and/or reverberations.

11 Ibid.

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

13 K.L. Payton and L.D. Braida, "Speech Modulation Transfer Functions for Different Speaking Style," *J. Acoust. Soc. Am.* 98: 2982 (1997).

14 K.L. Payton, R.M. Uchanski, and L.D. Braida, "Intelligibility of Conversational and Clear Speech in Noise and Reverberation for Listeners with Normal and Impaired Hearing," *J. Acoust. Soc. Am.* 95: 1581-92 (1994).

Our speech-based approach to computing the STI generally does not make accurate predictions of the differences in intelligibility between clear and conversational speech that are subject to the same degradations. Intelligibility has a different dependence on STI for clear speech than for conversational speech. However, this divergence is reduced for the speech-based STI, but not for the traditional STI, as the degradation becomes more severe. This implies that it may be possible to modify the speech-based STI to account for some of the intelligibility differences associated with speaking style in difficult listening conditions.

1.2.3 Publications

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

Payton, K.L., and L.D. Braida, "Speech Modulation Transfer Functions for Different Speaking Style." *J. Acoust. Soc. Am.* 98: 2982 (1997).

Payton, K.L., and L.D. Braida. "Determining the Speech Transmission Index Directly from Speech Waveforms." Submitted to *J. Acoust. Soc. Am.*

Thesis

Graf, I.J. *Simulation of the Effects of Sensorineural Hearing Loss.* M.Eng. thesis, Department of Electrical Engineering and Computer Science, September 1997.

1.3 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, Joseph A. Frisbie, David S. Lum, Sharieff A. Mansour, Matthew G. Sexton, Prasanna B. Tambe

1.3.1 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.

1.3.2 Studies and Results

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. Grant et al.[15] have shown that a single envelope signal (derived from an octave-band of speech centered at 500 Hz using a 50 Hz smoothing filter and presented by modulating a 500 or 200 Hz tone) can enhance speechreading substantially, providing a basis for face to face communication. Listeners with normal hearing can achieve higher levels of speech reception when a second envelope signal (derived from the octave band of speech centered at 3150 Hz) is included in the speechreading supplement. However, this benefit is greatest when the second envelope signal modulates a 3150 Hz tone; when tones of lower frequency are used (as would be necessary to aid listeners with severe hearing losses) the benefit provided by the second envelope decreases. In presenting multiple envelope signals, previous investigators tacitly assumed that the different amplitude modulated tones conveyed independent speechreading cues. However, the band envelope waveforms of different octaves of speech usually exhibit significant correlation. Based on our own experience, it seems more likely that the high-frequency tone is used to detect when the envelopes differ substantially (e.g., due to the occurrence of strong fricatives). As such it is not necessary to convey two independent envelope cues, but merely to signal the occurrence of discrepant envelopes. To explore this possibility, Mansour[16] studied the effects of seven speechreading supplements, including four based on monaural presentation and three based on binaural presentation. The coding schemes all presented the envelope of the octave band centered at 500 Hz by modulating a 200 Hz tone but differed in how the envelope of the octave band centered at 3300 Hz was conveyed. The supplements were

I. A 200 Hz tone modulated by the envelope of the 500 Hz octave of speech,

II. Supplement I to which a 500 Hz tone modulated by the envelope of the 3200 Hz octave of speech is added,

III. Supplement I to which a 240 Hz tone modulated by the envelope of the 500 Hz octave of speech is added,

IV. Supplement I to which a 240 Hz tone modulated by the envelope of the 3200 Hz octave of speech is added,

V. A 200 Hz tone modulated by the envelope of the 500 Hz octave of speech presented to one ear, a 500 Hz tone modulated by the envelope of the 3200 Hz octave of speech presented to the contralateral ear,

VI. A 200 Hz tone modulated by the envelope of the 500 Hz octave of speech presented to one ear, a 200 Hz tone modulated by the envelope of the 3200 Hz octave of speech presented to the contralateral ear,

VII. The sum of the two signals in supplement V presented to one ear, the difference of the signals presented to the opposite ear.

Supplements I and II were reference conditions that had been studied by Grant et al.[17] Supplement III consisted of a single band envelope signal that was conveyed by a "rough" carrier.[18] Supplement IV introduced a "roughness" percept when the two band envelopes were correlated. Supplement V, VI, and VII attempted to use dichotic presentation to minimize interference between the components of supplement II. Spectra of the acoustic signals corresponding to supplements II and IV are shown in Panel E.

Fifteen subjects with normal hearing were screened to identify four with good speechreading abilities: above 40% on words in CUNY[19] sentences and above 20% on words in low-context IEEE sentences.[20] Each subject participated in six two-hour training sessions using CUNY sentences and ten test sessions using IEEE sentences, with ten lists (500 keywords) used for each presentation condition. An analysis of variance indicated that both listener and condition, but not their interaction, had significant effects. Averaged over subjects, scores obtained in conditions I and III were similar (63-65%), as were scores in conditions II, IV, V, and VII (76, 74, 72, and

15 K.W. Grant, L.D. Braida, and R.J. Renn, "Auditory Supplements to Speechreading: Combining Amplitude Envelope Cues from Different Spectral Regions of Speech," *J. Acoust. Soc. Am.* 95: 1065-73 (1994).

16 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.

17 K.W. Grant, L.D. Braida, and R.J. Renn, "Auditory Supplements to Speechreading: Combining Amplitude Envelope Cues from Different Spectral Regions of Speech," *J. Acoust. Soc. Am.* 95: 1065-73 (1994).

18 E. Terhardt, "On the Perception of Periodic Sound Fluctuations (Roughness)," *Acustica* 30: 201-13 (1974).

19 A. Boothroyd, T. Hnath-Chisolmn, and L. Hanin. "A Sentence Test of Speech Perception: Reliability, Set Equivalence, and Short-Term Learning," Report No. RCI10 (New York: City University of New York, 1985).

20 Institute of Electrical and Electronics Engineers, *IEEE Recommended Practice for Speech Quality Measurements*, No. 297 (New York: IEEE, 1969).

73%). The average score in condition VI (68%) was intermediate between the others. Note that scores were highest, and essentially equal, for four supplements that conveyed two band-envelope signals. Surprisingly, monaural presentation led to scores that were slightly higher than binaural presentation. These results indicate that it is possible to increase the benefits provided by a single envelope signal by introducing a second envelope signal that elicits the sensation of roughness. The second envelope signal can thus be conveyed with only a small increase in bandwidth.

Supplements Based on Automatic Speech Recognition

In manual cued speech (MCS), a speaker produces hand gestures to resolve ambiguities among speech elements that are often confused by speechreaders. The shape of the hand distinguishes among consonants; the position of the hand relative to the face among vowels. Experienced receivers of MCS achieve nearly perfect reception of everyday connected speech. MCS has been taught to very young deaf children and greatly facilitates language learning, communication, and general education. However, cued speech is used by only a small proportion of deaf persons, most likely because few hearing persons are skilled at producing the cues and the availability of transliterators is limited.

To facilitate the use of cued speech, we are developing an automatic cueing system, ASCOT,[21] that uses automatic speech recognition (ASR) technology to derive cues from the acoustic speech waveform and display them on a video screen in synchrony with images of the face of the talker. The results of a simulation study[22] indicate that the phone recognition performance of current ASR systems can provide an adequate basis for an automatic cueing system. To operate successfully in real time, the automatic cueing system must recognize cues with high accuracy and display them with minimal delay relative to the talker's speech.

The ASR component of ASCOT acoustic speech waveform samples speech at 10kHz and segments the samples into 20 ms-long frames with 10 ms overlap. For each frame, a vector of 25 parameters is derived from the samples including 12 mel-frequency cepstral coefficients, 12 differences of cepstral coefficients across frames, and the difference between frame energies. Differences were computed over a four frame span (+/-20 ms) centered on the current frame. To improve robustness, we applied RASTA processing[23] to the parameter vectors.

Speech parameters are computed by a Motorola DSP96002 chip under the control of a Pentium PC. The Pentium provides time-marked speech parameter vectors to a DEC AlphaStation that performs phone recognition using code based on the HTK software.[24] Three-state Hidden-Markov phone models with parameters described by mixtures of six Gaussian densities were used after pilot studies found them to provide the highest accuracy for the available speech data (1100 sentences/talker). The HTK recognition code was modified to produce the running estimate of the spoken phones that was required for real-time operation. The recognized phones were converted into a sequence of cues via a finite-state grammar, together with estimates of start times.

The recognition process uses context-dependent phone models based on 13 phone classes. To achieve real-time operation, ASCOT uses a two-pass search technique similar to the Forward-Backward search algorithm.[25] In this technique, a set of phone sequence candidates is identified by a fast search that uses context-independent models. Context-dependent models are then used to identify the most probable sequence. The size of the set of candidate sequences is adjusted dynamically to maintain recognition speed. Measured phone accuracy was 79-81% in off-line tests and 72-74% in real time operation.

In manual cued speech, cues correspond to CV pairs. While the manual cuer can produce the cue at the start of the initial consonant, an automatic system

21 P. Duchnowski, D.S. Lum, J.C. Krause, M.G. Sexton, M.S. Bratakos, and L.D. Braida, "Development of Speechreading Supplements Based on Automatic Speech Recognition," submitted to *IEEE Trans. Biomed. Eng.*

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

23 H. Hermansky and N. Morgan, "RASTA Processing of Speech," *IEEE Trans. Acoust., Speech, Sig. Proc.* 34: 52-59 (1994).

24 P.C. Woodland, C.J. Leggetter, J.J. Odell, V. Valtchev, and S.J. Young, "The 1994 HTK Large Vocabulary Speech Recognition System," *Proceeding of the International Conference on Acoustics, Speech, and Signal Processing* 1: 73-76 (1995).

25 S. Austin, R. Schwartz, and P. Placeway, "The Forward-Backward Search Algorithm," *Proceedings of the 1991 International Conference on Speech, Acoustics, and Signal Processing* 1: 697-700 (1991).

cannot determine what cue to produce until after the final vowel is spoken and recognized. Typically this would impose a delay in the display of the cue that is well in excess of 165 ms, whose effect was found to be highly deleterious by Bratakos et al.[26] To compensate for this delay, the video image of the talker's face is stored for delayed playback when the corresponding cue is identified. Handshapes corresponding to the identified cues are superimposed on video frames of the talkers face at appropriate locations, and the dubbed images are displayed on a monitor.

ASCOT was evaluated by experienced cue receivers in speech reception tests using low-context[27] sentences materials. With the most advanced version of the ASCOT system, three subjects achieved word scores averaging 35% via speechreading alone, 90% via manual cued speech, and 66% when using cues produced by Ascot. Moreover, each of the three subjects commented that the ASCOT system provided them with palpable and significant aid relative to the speechreading alone condition. When compared to the speech reception scores of users of the Inneraid cochlear implant and the Tactaid 2 and Tactaid 7 tactile aids,[28] these results are encouraging. On comparable tests of speech reception, users of the ASCOT achieved word scores that exceed those of roughly one-third of cochlear implant users and two-thirds of those obtained with tactile aids. Although continued development of cochlear implants and tactile aids is likely to improve the effectiveness of these devices, continued development of ASCOT is clearly warranted.

Automatic Transcription of Recorded Speech

Our current automatic cueing system uses a speaker-dependent recognizer because speaker-independent speech recognizers are not sufficiently accurate to produce reliable cues. However, speaker-dependent recognizers require phone models that are trained on the utterances of a particular speaker. To achieve high accuracy, a 60-minute corpus of utterances must be phonetically transcribed and time-marked for each new speaker. Many weeks of effort by trained phoneticians are required to prepare the materials, thus limiting the number of speakers that can be used.

To expedite this process, we developed a software system that produces the required time-aligned phonetic transcriptions automatically.[29] The system uses the recorded speech of a new talker, the text (orthography) of the utterances, and speech models that have been trained on a multispeaker database, e.g., TIMIT.[30] In the first stage, the text of the utterances is used to produce a list of the possible phone sequences that could be used when the text is spoken. Three text-to-phone conversion sources are used to produce this list: a phonetic dictionary,[31] an algorithm that generates pronunciations for arbitrary letter sequences,[32] and a set of rules that we derived by analyzing databases that had been transcribed by phoneticians. The output of this stage is a phone network corresponding to the given utterance. In the second stage, the parameterized speech, the general phone models, and the phone network are aligned by a Viterbi search routine[33] to determine the phone sequence that best corresponds to the speech waveform. Because the number of possible phone sequences is highly constrained, the use of multi-speaker phone models is adequate to determine the phonetic transcription with an accuracy in excess of 90%, which is close to the degree of agreement between two human transcibers. These transcriptions can then be used to train the recognizers on the new speaker. We have also found these transcriptions to be useful in automating measurements of the acoustical properties of speech waveforms.

---

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

27 Institute of Electrical and Electronics Engineers, *IEEE Recommended Practice for Speech Quality Measurements*, No. 297 (New York: IEEE, 1969).

28 C.M. Reed and L.A. Delhorne, "Current Results of a Field Study of Adult Users of Tactile Aids," in *Tactile Aids for the Hearing Impaired,* ed. D.K. Oller. *Seminars in Hearing 16*, 1995, pp. 305-15.

29 S. Schlueter, *Automatic Refinement of Hidden Markov Models for Speech Recognition,* M.Eng. thesis, Department of Electrical Engineering and Computer Science, MIT, February 1997.

30 *TIMIT Acoustic-Phonetic Continuous Speech Corpus,* NIST Speech Disc 1-1.1, October 1990.

31 Carnegie Mellon University, *Carnegie Mellon Pronouncing Dictionary,* CMUdict.0.4, unrestricted on-line distribution, 1995.

32 J. Wasser, "English to Phoneme Translation," public software developed at the Digital Equipment Corporation, last revision April 15, 1985.

33 P.C. Woodland, C.J. Leggetter, J.J. Odell, V. Valtchev, and S.J. Young, "The 1994 HTK Large Vocabulary Speech Recognition System," *Proceeding of the International Conference on Acoustics, Speech and Signal Processing* 1: 73-76 (1995).

Measurement of Facial Actions

Characterizing the visible actions of the face that are used by speechreaders has proven challenging because these actions involve the three-dimensional motions and deformations of low-contrast tissue. To facilitate measurement of these actions, Tambe[34] has developed a system that uses image processing technique to track the motion of selected points on the surface of the face.

To facilitate the identification of the points, a high-contrast grid is painted on the skin surface and the grid is recovered from video images using adaptive thresholding. Active contours[35] such as snakes are used to track elements of the grid automatically. Snakes move under the effect of external forces corresponding to the intensity gradient in the image and internal forces provided by springs structured to maintain contour length and resist buckling of the contour. During facial actions, snakes assume configurations that minimize total energy. These configurations are determined via dynamic programming with initial estimates derived from configurations seen in prior frames. Three-dimensional coordinates of the selected points are derived from the spatial configurations of snakes seen in a pair of camera images using perspective projection.

1.3.3 Publications

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

Duchnowski, P., D.S. Lum, J.C. Krause, M.G. Sexton, M.S. Bratakos, and L.D. Braida, "Development of Speechreading Supplements Based on Automatic Speech Recognition." Submitted to *IEEE Trans. Biomed. Eng.*

Theses

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

Schlueter, S. *Automatic Refinement of Hidden Markov Models for Speech Recognition.* M.Eng. thesis, Department of Electrical Engineering and Computer Science, MIT, February 1997.

Sexton, M. *A Video Display System for an Automatic Cue Generator.* M.Eng. thesis, Department of Electrical Engineering and Computer Science, MIT, 1997.

Tambe, P.B. *Facial Feature Tracking using Deformable Templates.* M.Eng. thesis, Department of Electrical Engineering and Computer Science, MIT, May 1997.

1.4 Tactile Communication of Speech

Sponsor

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

Project Staff

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

1.4.1 Goals and Significance

This research is directed toward 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 these 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 an increased understanding of speech communication, environmental-sound reception, tactual perception, manual sensing, display design, and sensory substitution.

1.4.2 Studies and Results During 1997

Basic Studies of Hand Stimulation and Active Touch

Studies are being conducted to explore the capabilities of the tactual sense for conveying information. These studies employ a multifinger tactual display designed by Tan[36] to provide stimulation along a continuum from high-frequency, low-amplitude kines-

34 P.B. Tambe, *Facial Feature Tracking using Deformable Templates,* M.Eng. thesis, Department of Electrical Engineering and Computer Science, MIT, May 1997.

35 M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *Int. J. Contour Models* (1987).

thetic movements to low-amplitude, high-frequency cutaneous vibrations. Experiments currently underway are designed to: (1) provide further understanding of the properties of backward and forward masking on information transfer (IT); (2) determine the nature and structure of errors made in the presence of masking signals, and (3) provide a somewhat more direct approach towards estimating IT rates than that employed previously by Tan et al.[37]

Specifically, the goals of the current study are to: (1) measure the masking effects of motional, rough, and vibrational signals using typical forward- and backward-masking paradigms, (2) investigate the temporal integration properties of the signals employed in the stimulus set (e.g., a rough signal followed by a vibration may be perceived as the combination of both), and (3) establish the relationship between performance with simple masking paradigms and that with the identification paradigm used previously by Tan[38] for estimating information-transfer rate.

Current experiments are being conducted with a set of seven signals at two different durations (125 or 250 Hz) presented to the index finger of the left hand. 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 320 msec) and asked to identify either the second signal (in the forward-masking conditions) or the first signal (in the backward-masking conditions). In a third paradigm, subjects are presented with three successive stimuli (again using interstimulus intervals in the range of 0 to 320 msec) and asked to identify the middle signal or to identify all three signals.

Experimental results will be analyzed as a function of interstimulus duration at each of the two signal durations, and error trials will be examined to gain insight into the processes that mediate masking.

Evaluation of Wearable Tactile Aids

Research conducted during 1997 includes studies of speech reception and/or speech production in both children and adults fitted with wearable tactile aids. Work has continued with a small group of children at the Randolph Learning Center who are receiving speech instruction using tactile aids (specifically, the Tactaid 7 device). Results of this ongoing work are reported in Plant et al.[39] Step by Step, a speech training program aimed at teachers of the deaf, speech pathologists, and audiologists, stresses the use of tactile aids as a means of providing feedback for deaf speakers. Step by Step should be released and distributed in 1998 by the Hearing Rehabilitation Foundation. A modification to the Tactaid 7 device, in which five vibrators are used to convey high-frequency information in the region of 2-7 kHz, has been studied with a congenitally deaf adoloescent. This subject was provided with training in the use of the tactile display for both speech perception and production. Pre- and post-training recordings of the subject's speech productions were obtained for later analysis.

Development of Improved Tactual Supplements to Speechreading

We are continuing to examine the reception of voicing through a variety of tactual displays. Current research involves the ability to discriminate speech features through the electrotactile display of the Tickle Talker device, developed at the University of Melbourne. This device consists of an eight-channel array of electrodes worn on four fingers of the hand which are used to encode information about F0, F2, and speech amplitude through the electrical parameters of pulse rate, electrode position, and charge per pulse, respectively. Normal-hearing laboratory subjects are participating in a study of the discriminability of pairs of vowels and consonants as a function of token variability.

---

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

37 H.Z. Tan, N.I. Durlach, C.M. Reed, and W.M. Rabinowitz, "Information Transmission with a Multi-Finger Tactual Display," *Scandinavian Audiol.* 26, Suppl. 47: 24-28 (1997); H.Z. Tan, N.I. Durlach, C.M. Reed, and W.M. Rabinowitz, "Information Transmission with a Multi-Finger Tactual Display," *Percept. Psychophys.*, under revision.

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

39 G. Plant, M. Horan, and H. Reed, "Speech Teaching for Deaf Children in the Age of Bilingual/Bicultural Programs: The Role of Tactile Aids," *Scand. Audiol.* 26 (Suppl. 47): 19-23 (1997).

Study of the Reception of Environmental Sounds through Tactual Displays

A test of the ability to identify environmental sounds has been developed, employing closed sets of ten sounds in each of four different settings (office, outdoors, kitchen, and general home). Identification data are being collected on a group of profoundly deaf subjects who are regular users of the Tactaid 7 device. Testing is conducted using a one-interval, ten-alternative, forced-choice procedure using Matlab software.

Data, obtained first without feedback and then with trial-by-trial correct-answer feedback, are being analyzed in terms of overall percent correct and IT calculated from stimulus-response confusion matrices. Data currently analyzed indicate that subjects were able to identify roughly 45% of the sounds on the no-feedback conditions (where chance is 10%) and that their performance improved by roughly 20 percentage points when feedback was provided. Overall performance levels were similar across the four different environments. Thus, these results indicate that cues useful to the identification of environmental sounds, other than speech, are provided by the seven-channel spectral display of the Tactaid 7 device.

1.4.3 Publications

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

Reed, C.M., and N.I. Durlach. "Note on Information Transfer Rates in Human Communication." *PRESENCE.* Forthcoming.

Plant, G., M. Horan, and H. Reed. "Speech Teaching for Deaf Children in the Age of Bilingual/Bicultural Programs: The Role of Tactile Aids." *Scand. Audiol.* 26 (Suppl. 47): 19-23 (1997).

Plant, G. "Training in the Use of a Tactile Supplement to Lipreading: A Long-Term Case Study." *Ear Hear.* Forthcoming.

Tan, H.Z., N.I. Durlach, W.M. Rabinowitz, C.M. Reed, and J.R. Santos. "Reception of Morse Code through Motional, Vibrotactile, and Auditory Stimulation." *Percept. Psychophys.* 59: 1004-17 (1997).

Tan, H.Z., N.I. Durlach, C.M. Reed, and W.M. Rabinowitz. "Information Transmission with a Multi-Finger Tactual Display." *Percept. Psychophys.* Under revision.

Tan, H.Z., N.I. Durlach, C.M. Reed., and W.M. Rabinowitz. "Information Transmission with a Multi-Finger Tactual Display." *Scand. Audiol.* 26 (Suppl. 47): 24-28 (1997).

1.5 Multimicrophone Hearing Aids

Sponsor

National Institutes of Health
Grant 2 R01 DC00270

Project Staff

Joseph G. Desloge, Dr. Julie E. Greenberg, Michael P. O'Connell, Dr. William M. Rabinowitz, Dr. Patrick M. Zurek

The goal of this research is to determine how hearing aids can be improved 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.[40] Ideally, this intelligibility enhancement would be achieved without compromising the listener's ability to monitor and localize sound sources from all directions. Array processing algorithms are first implemented and evaluated in terms of signal-to-noise improvement in computer simulations. The most promising approaches are then implemented in real-time with wearable devices (tethered to a computer) for laboratory evaluations in terms of speech reception in noise and sound localization by normal and hearing-impaired listeners.

Recently-completed work focused on developing microphone array hearing aids with binaural outputs that allow users to localize sound sources.[41] Additional research efforts have modified existing adaptive algorithms to overcome problems caused by the

40 P.M. Zurek, J.E. Greenberg, and W.M. Rabinowitz, "Prospects and Limitations of Microphone-array Hearing Aids," in *Psychoacoustics, Speech, and Hearing Aids,* ed. B. Kollmeier (Singapore: World Scientific, 1995).

41 J.G. Desloge, W.M. Rabinowitz, and P.M. Zurek, "Microphone-array Hearing Aids with Binaural Output. I. Fixed-processing Systems," *IEEE Trans. Speech Audio Proc.* 5: 529-42 (1997); D.P. Welker, J.E. Greenberg, J.G. Desloge, and P.M. Zurek (1997). "Microphone-array Hearing Aids with Binaural Output. II. A Two-Microphone Adaptive System," *IEEE Trans. Speech Audio Proc.* 5: 543-51 (1997).

short-time power fluctuations in speech.[42] Current work is focused on a new adaptive algorithm that actively monitors and controls system behavior. This improved control offers enhanced robustness to adverse listening environments.

Publications

Desloge, J.G., W.M. Rabinowitz, and P.M. Zurek, "Microphone-array Hearing Aids with Binaural Output. I. Fixed-processing Systems," *IEEE Trans. Speech Audio Proc.* 5: 529-42 (1997).

Greenberg, J.E. "Modified LMS Algorithms for Speech Processing with an Adaptive Noise Canceller." *IEEE Trans. Speech Audio Proc.* Forthcoming.

Welker, D.P., J.E. Greenberg, J.G. Desloge, and P.M. Zurek. "Microphone-array Hearing Aids with Binaural Output. II. A Two-Microphone Adaptive System," *IEEE Trans. Speech Audio. Proc.* 5: 529-42 (1997).

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

A variety of 13 array-processing and reference algorithms have been compared on a battery of localization, speech reception, and pleasantness-rating tests.[43] Current work is aimed at a similar comparison of adaptive feedback cancellation algorithms.

Conference Paper

Greenberg, J.E., P.M. Zurek, W.M. Rabinowitz, M. Brantley, and A.R. Brughera. "Program of Hearing Aid Device Development: Algorithms." Second Biennial Hearing Aid Research and Development Conference, National Institutes of Health, Bethesda, Maryland, 1997.

1.7 Binaural Hearing

Sponsor

National Institutes of Health
Grant 2 R01 DC00100[44]

Project Staff

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

The long-term goal of this project is to (1) develop an integrated, quantitative theory of binaural interaction that is consistent with psychophysical and physiological data on normal and impaired auditory systems and (2) apply our results to the diagnosis and treatment of hearing impairments. Current work is in two areas: (1) study of the lateralization of transient noise bursts with time-varying interaural delays[45] and (2) analysis of the acoustic cues that underlie the localization of "phantom" sound images in intensity-difference stereophony.[46]

42 J.E. Greenberg, "Modified LMS Algorithms for Speech Processing with an Adaptive Noise Canceller," *IEEE Trans. Speech Audio Proc.*, forthcoming.

43 J.E. Greenberg, P.M. Zurek, W.M. Rabinowitz, M. Brantley, and A.R. Brughera, "Program of Hearing Aid Device Development: Algorithms," Second Biennial Hearing Aid Research and Development Conference, National Institutes of Health, Bethesda, Maryland, 1997.

44 Subcontract from Boston University. Professor H. Steven Colburn, Principal Investigator.

45 P.M. Zurek and K. Saberi, "Stimulus-token Effects in the Lateralization of Two-transient Stimuli," submitted to *J. Acoust. Soc. Am.*

46 P.M. Zurek and B.G. Shinn-Cunningham, "Localization Cues in Intensity-difference Stereophony," submitted to *J. Acoust. Soc. Am.*

Publications

Zurek, P.M., and K. Saberi, "Stimulus-token Effects in the Lateralization of Two-transient Stimuli," submitted to the *J. Acoust. Soc. Am.*

Zurek, P.M., and B.G. Shinn-Cunningham, "Localization Cues in Intensity-difference Stereophony," submitted to the *J. Acoust. Soc. Am.*

1.8 Virtual Environment Technology for Training

Sponsor

U.S. Navy - Office of Naval Research
Grant N61339-96-K-0002
Grant N61339-96-K-0003
Grant N00014-97-1-0635
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, David W. Schloerb, Adrienne H. Slaughter, Kari Anne H. Kjolaas, Lajos Molnar, Evan F. Wies, Chi-Hao Ho, Alexandra I. Hou, David E. DiFranco, Hanfeng Yuan, Wan-Chen Wu

1.8.1 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.

1.8.2 Visual Depth Perception in VEs

Research has continued on depth perception via stereopsis and motion parallax. The research on stereopsis, with emphasis on adaptation to increased interpupillary distance, has nearly been completed, with most of the results that we have obtained now available in David Schloerb's Ph.D. dissertation. Current work in this area is directed toward determining the effects of antialiasing on depth resolution and preparing articles for publication.

In an additional experiment related to stereopsis depth cues, we have explored the effect of these cues on contrast induction. Contrast induction is an example of lateral interaction, a form of interaction that is common in visual perception; perceived brightness as well as chromaticity of a region in a scene are affected by nearby regions, as is apparent contrast and the perception of motion. These effects can be regarded as a means to augment differences between regions and may serve to enhance discriminability.

The availability of stereoscopic displays for work in VEs raises the issue whether stereo depth will affect induction processes and possibly discriminability—factors which may need to be considered in the design of VE display systems. For example, depth layout of a scene has been shown to affect perceived brightness.

In this set of experiments, we began to explore the effects of stereoscopic depth cues on contrast induction, in which the perceived contrast of a textured test region is affected by the contrast of a textured surround. A central test region was surrounded by an annular inducing region. Both regions were filled with random-dot patterns whose contrast could be modulated in time. In order to reduce inhomogeneous local brightness adaptation, which may lead to a contamination of the stimulus with brightness induction, a different random dot pattern was presented for each frame of the display, which was updated at 60 Hz.

This resulted in a dynamic random noise stimulus in which the time-averaged mean luminance was the same for all points.

Stimuli were viewed through Crystal Eyes stereo glasses; in this way, the disparity of the test and surround regions could be made to differ. Though the stimulus had the appearance of "snow" on a television set due to its dynamic random nature, the correlation of dots between left and right eyes elicited a stable depth percept. This stimulus configuration could be used to simulate two surfaces at different depths onto which dynamic dot patterns were projected.

The magnitude of induced contrast was estimated for three test conditions: in the first, the test and surround both lay in the plane of the screen. For the other two, the disparity of the surround was modified so that the annulus appeared to either lie in front of the plane of the screen, and thus in front of the test or behind it. A nulling method, was used to estimate the magnitude of induced contrast in the test region. Modulating the contrast of the surround sinusoidally in time induced a modulation in the perceived contrast of the test region which was in opposite phase to that of the surround. This induced modulation could be nulled with a real contrast modulation inside of test field. The magnitude of the modulation required to null the apparent modulation of the test provided an estimate of the magnitude of induced contrast.

Because observers find it difficult to precisely set the nulling amplitude to achieve an exact null, the method of constant stimuli was used. A short tone was synchronized with the trough of the inducing modulation, while a longer tone indicated the peak of the inducing modulation. The observer indicated by pressing a button whether the perceived contrast of the test region was higher at the phase of the modulation indicated by the long tone or vice versa. The percentage of responses corresponding to a perceived high contrast during the long tone was plotted as a function of nulling amplitude to yield psychometric curves for the nulling of induced contrast.

The resulting psychometric functions (obtained for three observers and fitted with Weibull curves) show that a lesser nulling modulation was required when the test and surround regions lie in different depth planes than when they were coplanar. These results suggest that 3D information in visual scenes may affect perceived contrast phenomena.

Research on depth perception via motion parallax has been seriously hampered by the departure of the postdoctoral fellow who was working on this project and by difficulties with the experimental apparatus. During the past year we have addressed the critical problems with the apparatus and are now able to present headtracked, antialiased, random dot stimuli, rendered at 60 Hz, which convey depth solely through motion parallax cues.

We have decided to temporarily forgo the use of the servo-controlled "bite bar" because of a variety of ongoing problems with both the mechanism and the control algorithms. The major result which we sought in using this complex arrangement was precise control of the velocity of the users head motion. For the near-term experiments, control of velocity is achieved through subject instruction coupled with elimination of trials in which the measured velocity exceeds an acceptable range.

In our earlier experiments, subjects viewed a split-field random dot display, and their ability to identify the nearer field was measured. We have since completed a number of enhancements to this presentation. Improvements in the antialiasing procedure used for rendering the dots, yielding finer spatial resolution and smoother motion, was implemented. Also, the split-field display was reprogrammed to instead follow a sinusoidal curve in depth. (The previous split-field arrangement could be thought of as a square wave in depth, the high spatial frequency components at the depth transitions were potential artifacts that have been eliminated in the current arrangement.)

The psychometric procedure (interleaved staircases) is in place to determine thresholds for the perception of depth in these stimuli as we sample the parameter space consisting initially of the following three dimensions: (1) frame rate, (2) spatial resolution of head position information, and (3) delay between head motion and stimulus display. The experimental conditions are organized along these three dimensions, and the resulting threshold values for each combination of the conditions will reveal the areas of greatest perceptual sensitivity.

A second set of experiments based on these findings has also been designed to explore the effects of spatiotemporal jitter (uncertainty in head position measurement), and the effects of spatial quantization in the display (pixel size and resulting voxel size), measures which have particular relevance with respect to limits on acceptable inter-process delays and periodicity, and the required resolution of the display system.

We have also produced a collection of random dot displays of three-dimensional objects which can be disambiguated through motion parallax. Initial experience with these stimuli confirms previous anecdotal reports of a long time-constant for the appearance and subsequent decay of the depth effect. The existence of this slow process may provide insight about the level of sensitivity of the depth percept to spatiotemporal jitter. Systematic exploration of these results may be warranted, pending the results of the sinusoidal curve experiments.

1.8.3 Sensorimotor Loop Research on Adaptation to Altered Relations Between Visual and Proprioceptive/ Kinesthetic Information on Hand Position

During the past funding period, we completed a long series of experiments on adaptation to alterations based on simple rectilinear and polar transformations. These experiments were then augmented by experiments in which the transformation was more difficult for subjects to conceptualize in terms of simple coordinate transformations.

In order to examine the many relevant aspects of the resulting data, a collection of processing macros had to be written (in Excel). This task turned out to be more labor-intensive than expected; however it did enable us to process the large amounts of data in a manner that is now facilitating model development.

The current approach to modeling these results continues to center on a two-component adaptation process. The first component is assumed to be directed toward compensation for a simple approximation to the imposed transformation and to have a relatively fast adaptation rate; the second process is directed toward refinement of the approximation and to have a relatively slow adaptation rate (i.e., its effect is not noticeable until the first is essentially complete). The most crucial problems faced in the development of this model concern the construction of adequate general criteria for defining the characteristics of the approximations used in the first process (i.e., for defining "simplicity"). In developing this model, attention is being given not only to the above described data, but also to other data that have been collected on sensorimotor adaptation in our laboratory (e.g., concerning auditory localization), as well as results available in the literature.

1.8.4 Sensorimotor Involvement for the Training of Cognitive Skills: Haptic Enhancement of Memory

Previous experiments examining the effects of haptic involvement in the memorization and recall of simple sequences (see section 3.5 in our ERHO proposal from last year) produced large intersubject variability and unclear results. However, there was no indication that the addition of haptic involvement improved performance over the performance achieved with vision alone.

In an attempt to clarify this issue, we modified the experimental procedure in a number of ways. First, all experiments were administered in a forced choice paradigm. Second, the stimulus sets were tested separately rather than intermixed. Third, the response mode was varied.

Experiment 4.0 was designed to determine whether the manner in which the conditions are tested (i.e., sets of sequences tested separately as opposed to intermixed) has an effect on performance. Subjects included two groups of 12 subjects each. The stimuli were two sets of sequences (set 1 and set 2) tested in two modes (visual versus visual/haptic). All testing was forced choice. Training and testing for each group occurred over three sessions (three consecutive days).

Group 1 was trained on set 1 in the visual mode. Testing occurred at 5 minutes, 30 minutes and 24 hours after training and consisted of writing the sequences. They were then trained on Set 2 in the visual/haptic mode and tested 5 minutes, 30 minutes and 24 hours after training. Again, testing consisted of writing the sequences. At the end of the experiment, the stimulus sets were intermixed and the subjects tested again.

Group 2 was trained and tested in the same way but the order of presentation was reversed. In other words, group 2 received the visual/haptic condition first followed by the visual mode. The results of this experiment indicated that haptic involvement during training did not necessarily improve performance. The subjects in both groups received higher scores on the condition that they received last. In other words, group 1 showed higher performance in the visual/haptic mode and group 2 showed higher performance in the visual mode. The results also indicated that the manner of testing did not matter: subjects scored as well on the intermixed test condition as they did when tested on each stimulus set separately.

Another result that emerged from this experiment, which confirmed a surprising result obtained in our previous experiments, was that memory for test sets did not decrease as a function of time: subjects did as well when tested 24 hours after training as they did when tested five minutes after training.

At the end of the final test session, all subjects were asked for subjective impressions of the experiment by filling out a questionnaire and through a personal interview with the experimenter. One intersubject variable that was revealed by these subjective responses was that although the matrix was always available during training in the visual mode, the degree to which a subject used or relied on the matrix varied among subjects. Another problem that occurred during the experiment was that some subjects reached, or nearly reached, ceiling performance. To eliminate the issue of how much the matrix is utilized in a particular subject's training/testing, we made the decision to uniformly instruct all subjects on the use of the matrix in future experiments. To reduce the ceiling effect, the stimulus set size was increased to make future experiments more difficult. Finally, we eliminated the tests that occurred 30 minutes and 24 hours after training and retained only the five-minute post-training test.

Experiment 5.0 was designed to explore the effect of response mode. Do subjects, as expected, score better when the training and testing are done in the same mode? Six subjects were tested under the following 3 conditions. Condition 1: training = visual; testing = writes letter/number sequence; condition 2: training = haptic; testing = writes letter/number sequentially; condition 3: training = haptic; testing = draws line on matrix. In an attempt to eliminate ceiling effects, the stimulus set sizes were increased from four sequences to six sequences. All subjects received the conditions in the same order (condition 1, 2, 3).

Saturation levels were still present making it difficult to determine the effect of response mode. Therefore, the set size was increased from six sequences to eight sequences, and the experiment was repeated with five new subjects. Again near saturation levels were seen for all but one subject. However, the one subject who did not approach ceiling, showed a systematic increase in performance with each condition. In order to see if there was a learning effect associated with mastering the task, we tested her with new sets of stimuli in reverse order (conditions 3, 2, 1). Her performance was similar across conditions and was similar to her performance in the last condition (condition 3) of the previous experiment. We interpreted these findings as evidence of a learning effect in the previous experiment and that once plateau performance was reached there was no difference among conditions. We are in the process of repeating experiment 5.0 (see experiment 6.0 below) with some modifications. The test stimuli will be more difficult/complex and a repeat condition will be added to examine learning effects. Because of this repeat condition, condition 2 (training = haptic; testing = writes letter/number sequentially) will be eliminated to enable completion of testing within one session.

Experiment 6.0 will be conducted with two groups of subjects in two conditions. (The only difference between the groups will be the order of presentation.) The size of the stimulus set remains at 8 sequences but the size of the matrix is increased to 4 x 4 (from 3 x 3). The two conditions are condition 1: training = visual; testing = writes letter/number sequence and condition 2: training = haptic; testing = draws line on matrix. Whichever condition is presented first will be repeated at the end (with a different set of stimuli) to test for learning effects. In other words, Group 1 will receive the training/testing in the following order: condition 1, 2, 1. For group 2, the order of training/testing will be reversed (condition 2, 1, 2).

1.8.5 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 has opened up novel and interesting research areas.

Over the past few years, we have developed device hardware, interaction software and psychophysical experiments pertaining to haptic interactions with virtual environments (recent reviews on haptic interfaces can be found in Srinivasan 1995[47] and Srinivasan and Basdogan 1997[48]). 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. Kenneth Salisbury's group at the MIT Artificial Intelligence Laboratory, has been used to prototype a wide range of force-based haptic display primitives. A variety of haptic rendering algorithms for displaying the shape, compliance, texture, and friction of solid surfaces have been implemented on the PHANToM. All the three devices have been used to perform psychophysical experiments aimed at characterizing the sensorimotor abilities of the human user and the effectiveness of computationally efficient rendering algorithms in conveying the desired object properties to the human user.

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

1.8.6 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 sub-discipline 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 a hierarchical database, multi-threading 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 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 end point is inside the virtual object.[49] 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.[50] Both techniques have advantages and disadvantages. For example, it is computationally less expensive to render 3D objects using 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.

1.8.7 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 process 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 differ-

47 M.A. Srinivasan, "Haptic Interfaces," in *Virtual Reality: Scientific and Technical Challenges*, eds. N.I. Durlach and A.S. Mavor, pp. 161-87 (National Academy Press, 1995).

48 M.A. Srinivasan, C. Basdogan, "Haptics in Virtual Environments: Taxonomy, Research Status, and Challenges", *Computers and Graphics* 21(4): 393-404 (1997).

49 C. Ho, C. Basdogan, M.A. Srinivasan, "Haptic Rendering: Point- and Ray-based Interactions," *Proceedings of the Second PHANToM Users Group Workshop*, Dedham, Massachusetts, October 1997.

50 C. Basdogan, C. Ho, M.A. Srinivasan, "A Ray-based Haptic Rendering Technique for Displaying Shape and Texture of 3D Objects in Virtual Environments," Winter Annual Meeting of ASME '97, *DSC* 61:77-84, Dallas, Texas, November 1997.

ent 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.

1.8.8 Generating Multimodal Stimuli

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

1.8.9 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 details. 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. With this additional information, we are now able to render macro textures (> 5.0 mm height) as well.

We have also experimented with 2D 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 2D 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 2D images of spots (i.e., simple 2D 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 which are solved in advance to generate a texture field that can be mapped onto the 3D surface of the object), and (5) spot-noise textures (the noise function is convolved with 2D images of spots to generate distorted spots that can be displayed haptically). In addition, we have developed image-based haptic textures; the gray scale values of an image are used to generate texture fields that can be mapped onto the surface of 3D 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 play-back 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 play-back of virtual textures.

1.8.10 Compliance

We have developed procedures for simulating compliant objects in virtual environments. The developed algorithms deal directly with geometry of 3D 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.

1.8.11 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 not only provides 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.[51] However, the influence of contact sounds on the perception of object stiffness is not as dramatic when tapping virtual objects through a haptic interface.[52] 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.

51 M.A. Srinivasan, G.L. Beauregard, 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* 58: 555-59, Atlanta, Georgia, 1996.

52 D.E. DiFranco, G.L. Beauregard, 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* 61: 17-22, Dallas, Texas.

1.8.12 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 humans discriminate stiffness, viscosity or mass 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 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.

1.8.13 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 three 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.

Discrimination Experiment

In the discrimination experiment, subjects were presented with two virtual surfaces of different stiffnesses (stiffness difference of 50%). Surface 1 (K = 0.5 N/mm) and surface 2 (K = 0.75 N/mm) could either be presented as the left or the right surface. Subjects were asked to tap the surfaces and judge which one is stiffer. They were allowed to tap each of the surfaces as many times as they wished. As they tapped the surface, they would hear the same contact sound that could be unknown to them at the time of contact, before the time of contact, or after the time of contact. During each trial, one surface had an at-contact sound and the other surface could have a pre-contact, at-contact, or post-contact sound. Based on our original hypothesis, the pre-contact and post-contact sounds were always associated with more stiff and less stiff surfaces, respectively. The pre-contact and post-contact sound could be presented 0.5 mm, 1 mm, 1.5 mm, or 2 mm away from the contact surface. In four sessions, the subjects were presented with a total of 40 trials for each variation in a random order. A total of ten subjects, six female and four male, participated in the experiment.

The results show that in trials when both the stiffnesses had at-contact sounds, the subjects were almost 95% correct, indicating that the stiffness difference was easily discriminable. However, the stimuli with pre-contact sounds were judged to be less stiff than the ones with at-contact sound, even though the latter were haptically 50% less stiff. More over, this perceptual illusion caused the subjects' percentage correct calls to drop to as low as 30%, with the amount of drop roughly proportional to the distance corresponding to the location of the pre-contact sound. An analysis of variance of the results shows that the percentage correct for pre-contact sounds at 2 mm, 1.5 mm, and 1.0 mm are significantly lower than the percentage correct for at-contact sound ($P < 0.001$). This indicates that a pre-contact sound located at a distance of about 1 mm away from the contact surface will cause a perceptual illusion that the surface tapped is equal in stiffness to the one that is 50% less stiff, but has an identical at-contact sound.

In contrast, in trials with post-contact sound stimuli, the subjects' performance was at levels higher than 95% correct. This result indicates that the subjects perceived the stimuli with the post-contact sounds to be less stiff than those with at-contact sounds, which in fact had higher haptic stiffness. This result contradicted our initial hypothesis, but from these results it was unclear whether the subjects ignored the post-contact sounds and judged purely on the basis of haptic stiffness or whether the post-contact sounds also cause the perceived stiffness to be less, as in the case of pre-contact sound stimuli. We therefore conducted the following matching experiment to resolve this issue.

Matching Experiment

In the matching experiment, the subjects were again presented with two virtual surfaces. The stiffness of the left surface remained the same throughout the session with a value of $K = 0.75$ N/mm and an at-contact sound. The right surface had one of the 9 pre-, at-, or post-contact sound presentation variations and a random initial stiffness value, ranging from 0.25 N/mm to 1.25 N/mm. The subjects were asked to change the stiffness value of the right surface by increasing or decreasing its value (in 0.05 N/mm increments) by using a key until they perceived the two surfaces to be of equal stiffness. As in the discrimination experiments, the subjects were allowed to tap the surfaces as many times as they wished. In each of two sessions, there were eleven trials of each of the nine variations presented in a random order. The initial stiffness of the right surface had the values of 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95, 1.05, 1.15, or 1.25 N/mm. A total of seven subjects, four female and three male, participated in this experiment.

The results of the matching experiments show that the subjects were able to accurately match the stiffness when given the at-contact sound (0.76 ± 0.017 N/mm) to the reference stiffness (0.75 N/mm). When asked to match the stiffness of a surface that had a pre-contact or post-contact sound, subjects matched a stiffness that is significantly stiffer ($P < 0.001$) than the reference for pre-contact sounds at 2 mm, 1.5 mm, 1 mm, and post-contact sounds at 2 mm. As the pre-contact or post-contact distance increased, so did the matched stiffness, ranging from 0.76 N/mm (i.e., the matched stiffness with at-contact sound) to about 1 N/mm. This indicates that both pre- and post-contact sounds cause the perceived stiffness to be less than the same haptic stiffness presented with at-contact sound. Therefore, in order to match the reference stiffness with at-contact sound, the subjects overcompensated the haptic stiffness that was under their control. This interpretation is consistent with the results of the discrimination experiments. Further, the results indicate that the subjects do not ignore the post-contact sounds, a possibility raised by the discrimination experiment.

1.8.14 Haptic-visual Interactions

Previously we have shown how the perception of haptic stiffness is influenced by the visual display of object deformation.[53] 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.

---

53 M.A. Srinivasan, G.L. Beauregard, 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* 58: 555-59, Atlanta, Georgia, 1996.

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 3D objects. Due to perspective graphics, the objects that are far away are represented by smaller visual images. We investigated if they will feel smaller when their size and dimensions are explored using a haptic device. In the first set of experiments, a 3D rectangular block with two grooves on it (one is located behind the other) was displayed. For each trial, the size and the dimensions of the groves were modified relative to each other and the subjects were asked to differentiate them. A similar experiment was conducted with two buttons instead of groves to study the effect of perspective on perceptual feeling of compliance. In the second set of experiments, we graphically and haptically displayed two buttons (one located behind the other) and asked the subjects to differentiate the stiffness of the buttons. Our preliminary studies indicate that subjects feel the distant objects to be haptically smaller and stiffer.

1.8.15 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. dimensions of the mechanical links of the device for effective force feedback and the reachable distances 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. Currently, we are testing the feasibility of these approaches for integration into our existing multimodal software.

1.8.16 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.

1.8.17 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 3D 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 3D objects in the scene. Since our ray-based rendering technique relies on detecting collisions between a line segment model of the stylus and the 3D 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 3D 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.

1.8.18 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 in Section 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. The application program has been placed on the Laboratory of Human and Machine Haptics web page (http://touchlab.mit.edu), to be executed by any remote user with a PHANToM and a Windows NT computer running Netscape for WWW access. Remote users can download a dynamic link library and some Java classes from the web page to their computer, and then run the program in their web browser. Users are asked to discriminate the stiffness of sets of two springs, displayed visually on the screen and haptically with the PHANToM, and to send in their responses via an e-mail window in the web page. Thus, we now have the ability to perform perceptual experiments with multimodal VEs across the internet. Over the past year, the industry-standard virtual reality modeling language (VRML version

2.0) was extended to accommodate the haptic modality, allowing the rapid prototyping and development of multi-modal applications.

1.9 Training for Remote Sensing and Manipulation

Sponsor

U.S. Navy - Office of Naval Research
Subcontract 40167

Project Staff

Nathaniel I. Durlach, Dr. Thomas E.v. Wiegand, Dorrie Hall, David W. Schloerb, Andrew R. Brughera, Rebecca L. Garnett, David H. Manowitz, Brian D. Kincy, Jonathan L. Zalesky, Matthew E. Esch

1.9.1 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 *Progress Report* No. 139.[54] MIT's work on the TRANSOM project during the past year is described in the following subsections.

1.9.2 Collision Detection

One major component of any realistic operation within a virtual environment consists of the ability of the VE program to detect and report collisions between virtual objects. For the requirements of the TRANSOM project, specific information about vehicle behavior under conditions of collision needs to be made available to the operator, independent of whether the collisions occur between the ROV and boundaries of the undersea environment (including the sea floor), between the ROV and potential target objects, or between the ROV and its own tether. As reported at the end of the first year, collision detection with the sea floor and with the exterior boundaries of the undersea environment was made operational as a first step, prior to the development of the more complex algorithms that would enable collision detection to operate with respect to all other objects in the virtual world.

Implementation of a collision detection algorithm for arbitrary objects in the simulation was completed prior to the expert pilot experiments scheduled in February 1997, but an error in implementation was discovered; this was corrected in March. Although a significant latency in the frame rates of the VE software was discovered during this time, the collision detection module was not at first implicated as a causal factor. However, improvements in the management of the Inventor database did provide dramatic increases in performance in the simulation. Modifications to the Inventor models were also introduced both to increase performance and to prevent the memory leaks which apparently occurred due to the use of incompatible Inventor format versions. These two changes greatly reduced the amount of time spent in object-to-object collision detection routines and had a dramatic overall effect on the simulation update rate. The addition of greater detail in the objects checked for collisions yielded results which made the simulation seem increasingly realistic.

The collision detection in the TRANSoM testbed system was implemented as a set of Scheme procedures, run during the main simulation loop. This system tests for collisions using a hierarchical bounding box test. Given two Inventor scene graphs for objects, first bounding boxes are computed for both objects, and these are tested for collisions. If no collisions occur, then the procedure ends. If a collision is detected and the node being checked is a group node, then a further (recursive) check is performed on each of its children until no further collision is detected. These further checks are performed to ensure that the bounding box collision check is not really finding a collision between parts of the bounding boxes not occupied by the models being checked. Additionally, the recursive checks can be ended by adding one of two labels to a scene graph. The first such label, "ignore," tells the collision checking not to check any part of the scene graph below the label. The second such label, "bounding box", tells the collision checking to only check the next subtree for a bounding box collision and to stop recursing at this level.

This checking algorithm should give us reasonable detection of collisions. However, arbitrary object-aligned bounding box intersection testing is not supported in Open Inventor. As a result, we only approximate these checks, using a point near the front of

54 N.I. Durlach et. al., "Training for Remote Sensing and Manipulation," *MIT RLE Progress Report* 139: 388-95 (1996).

the ROV to approximate the ROV's location, and this point is tested for intersection with another model's bounding box. We are currently trying to implement this algorithm in full, as described above, since it will give quick results when no collisions occur (the normal case), but should not take inordinate amounts of time to detect a collision when it does occur.

1.9.3 Tether Projects

It is well known that some of the main difficulties in piloting ROVs result from the existence of the tether. Even when no entanglement occurs, the tether alters the dynamics of the vehicle. When an entanglement occurs, the effects can be severe, both in terms of destroying the mobility of the ROV and in the difficulty involved in untangling the tether.

In our research directed toward solving the tether problem, we have initiated three projects, each of which is briefly described in the following three subsections.

Development of a Tether Information System (TIS)

The goal of this project is to design and construct a system that provides information on the time-varying geometry of the tether (3D position of all points on the tether) and on the time-varying force exerted by the tether at the ROV attachment point as functions of (1) the physical specifications of the tether, (2) the movements of the ROV to which the tether is attached, (3) the time-varying length of the tether as determined by the tether supply management, (4) the external forces (such as currents) acting on the tether, and (5) signals derived from sensors imbedded in the tether. Although it might also be useful to include in the TIS some information concerning the positional relationship between the tether and other objects in the environment (e.g., by means of proximity sensors), consideration of such relational information is being postponed until later in the program. Also, forces internal to the tether are being considered only to the extent that they influence tether geometry.

At one extreme, a TIS can be based solely on a theoretical physical model of tether behavior. At the other extreme, the information provided at the output of the TIS can be based solely on signals from sensors in the tether. The task at hand is to choose a combination of model and sensor system that maximizes the accuracy of the tether information with minimum cost, where cost is defined in terms of dollars, complexity, etc.

Initial work on this tether project is concerned with TIS design. This work includes the (1) study of various tether models; (2) study of available (or constructible) sensors for possible deployment in the tether; (3) study of the tether information that can be obtained from hypothesized sensor arrays as a function of the performance characteristics of the sensors, the number and placement of these sensors along the tether, and the physical characteristics of the tether; and (4) estimation of the TIS output for various (tether model) x (sensor array) combinations.

Work to date on this project has focused on the study of tether models. A brief description of the models considered is presented in the following paragraphs.

In the models initially studied, the situation is static: the ROV position and tether length are assumed constant; the tether geometry and tension are determined solely by the positions of the endpoints of the tether and the water-current vector; and the tether geometry and tension are described only in the steady state (i.e., after the current has been acting on the tether for a long time period). In Yoerger's model, discussed further below, the only forces considered are the drag forces due to the current and the tension in the tether; the effects of depth and buoyancy are ignored; and the effects of stretching are ignored. A finite element approach is used and the solution involves an iterative procedure that is computationally expensive and fails when there is no current. In the models considered by Dennis,[55] funicular curves are used. Although these models have not yet been implemented, they appear to be much more efficient computationally and to be more easily applied to 3D than that of Yoerger. However, they suffer from the same limitations with respect to the physical parameters considered and the small-current problem.

The dynamic models being considered include Ablow and Schecter's 3D high-tension model,[56] Howell's 2D low-tension model,[57] and Hover's 3D massless

55 N. Dennis, "The 'Two-century Plus History of the Constant-tension Catenary as Applied to Tall Ship Sails, Paravan and Other Tows, Oil Slick Catch Booms and Some Buoy Moors," *Oceans '95 MTS/IEEE Conference Proceedings*, San Diego, California, October 9-12, 1995; N. Dennis, "On the Formation of Funicular Curves," *Int. J. Mech. Sci.* 36(3): 183-88 (1994).

model.[58] The 3D high-tension model takes account of vehicle movement and includes a tether stretching term; however, it is not applicable to cases in which the tension is zero somewhere along the tether. The Howell 2D low-tension model assumes the cable is inelastic but includes a bending stiffness term. This model appears to be applicable when the current is near zero and/or the tether is neutrally buoyant. In Hover's model, a stretching term is included and an attempt is made to reduce the computation time required in Howell's model by assuming that the tether is massless. Although this model promises to reduce computation time, its accuracy has not yet been adequately studied.

To date, Yoerger's is the only model that has been implemented. The inputs to this model are the endpoints of the tether, the current vector, the number of tether segments to be used in the finite element analysis, and a number of tether parameters (length, diameter, normal and tangential drag coefficients). The outputs of the model are the coordinates of the tether segments and the force vector acting on the ROV at the point of attachment. The program derives a numerical solution using an iterative process and a Nelder-Mead Simplex method to solve the numerical problem in each iteration. Apart from the computational expense of this method and the need for nonzero current for the method to be applicable, the method fails when the current direction coincides with the direction between the ROV and the home base (i.e., there is no normal force acting on the tether). Attention is now being turned in this physical modeling work to implementation of dynamic models.

Development of the Tether Component of the Integrated Virtual Environment Testbed

The goal of this project is to extend the VE training testbed to include a tether model and tether display that can serve as an effective training aid. With respect to the tether model, this objective implies an emphasis on (1) the problem of entanglements and (2) real-time simulation of tether behavior. Accordingly, current research on the tether model in this project is focused on tether geometry (as opposed to the effect of the tether force vector on ROV dynamics), finding simple approximations to the tether behavior geometry generated by the physical tether models being developed in the project described in section 3.1.1, and on finding simple means for simulating entanglements.

Simulation of entanglements requires collision detection (of the tether with objects in the environment as well as with itself) and modeling of the effects of the collision. The simplest model for the effects of a collision (which is the most conservative from a training viewpoint) results from assuming that the tether sticks to any object it touches without any slippage. More realistic models being considered assume the sticking only occurs when certain angular constraints (related to the extent to which the tether direction changes at the point of contact) are exceeded. Although further realism can be introduced through the introduction of friction/slippage parameters, it seems unlikely at present that the benefits of adding such complexity will outweigh the drawbacks.

Initial software work on this project has been directed toward the development of a tether object interface and API that will allow all current tether-related projects to proceed from a common code base and permit us to easily modify various implementations as they are developed.

The display component of the tether system will provide the link between the tether model and the rest of the virtual environment. As such, it will use the tether API to obtain coordinates of various points on the tether, and then use those coordinates to make a graphical tether object. Since graphical collision detection methods have been the subject of much research, the graphical object can be tested for collisions with other objects in the VE. Using these results, the display system will update the status of the tether model with respect to collisions, and the model will change behavior accordingly. The display system may also include other features such as collision indicators, but these features will only be needed after the main features of the system have been implemented.

The initial tether API has been specified as a C++-style object without any external reference to a particular tether model to be used so that various mod-

56 C.M. Ablow and S. Schecter, "Numerical Simulation of Undersea Cable Dynamics," *Ocean Eng.* 10(6): 443-57 (1983).

57 M.A. Grosenbaugh, C.T. Howell and S. Moxnes, "Simulating the Dynamics of Underwater Vehicles with Low-tension Tethers," *Int. J. Offshore Polar Eng.* 3(3): 213-18 (1993); C.T. Howell, "Numerical Analysis of 2-D Nonlinear Cable Equations with Applications to Low-tension Problems," *Int. J. Offshore Polar Eng.* 2(2): 110-13 (1992).

58 A. Blieke, *Dynamic Analysis of Single Span Cables*, Ph.D. diss., Department of Ocean Engineering, MIT, 1984.

els may be tried without affecting any external code. The API has general accesser and modifier methods for setting parameters, along with methods for obtaining tether coordinates either as arrays or one-at-a-time, and methods for modifying the state of the tether with respect to collisions. The model has been specified to allow it to modify the representation of the tether upon being notified of a collision. This feature is most noticeable when obtaining tether coordinates as arrays. As currently specified, the coordinates will be returned as arrays of segments, where each segment is a part of the tether between two collision points or ends of the tether. By making the tether model general in this regard, either the model itself or external code can handle the various pieces of the tether upon a collision.

Development of a PC-Based Part-Task Tether Trainer

The objective of this project is to develop a highly simplified, special purpose, experimental part-task training system. We can explore in a highly focused manner how tether displays can be used to train subjects to estimate tether geometry based solely on information about the physical properties of the tether, the environmental conditions, the ROV motion history, and the tether supply history. As in the project described in Sec. 3.1.2, the model of tether behavior will make use of approximations that allow real-time simulation.

Illustrative examples of the training exercises being considered include (1) training to estimate tether position and (2) training to avoid tether entanglements and inappropriate tether supply. In the first case, the stimulus on each trial consists of a water-current vector, an ROV path history, and a tether supply history; the response consists of an estimate of tether position; and the feedback consists of presenting the subject with the correct tether position (according to the tether model used). In the second case, the subject performs a "go and fetch" task in a cluttered environment (with currents) using a joystick to control the ROV and a slide to control the tether supply. The subject is instructed to perform the task in minimum time with minimum entanglements and without running out of tether. A payoff matrix is used to specify the penalties for entanglements, running out of tether, and time consumed in performing the task.

In initial work, both the ROV movement and the tether configuration are being restricted to a single plane, and the subject's viewpoint is normal to this plane. Later on, a viewpoint from inside the ROV will be added. Code for the VE is being written in C++ for a Pentium-based system running Microsoft Windows 95.

1.9.4 Spatialized Audio Displays

The objective of this project is to develop and evaluate various approaches to the display of spatialized auditory information for ultimate inclusion in the TRANSoM trainer system. Because of the many attentional demands involved in the task of piloting an ROV, it would appear to be advantageous to utilize all sense modalities when conveying information to the pilot. To this end, we are exploring ways in which "continuous" information, such as the state of the ROV thrusters, can be displayed as a spatialized sound field. Such a display, if successful, would constitute a separate channel of information which is fairly independent of the "discrete" (speech-based) channel presently used to provide coaching and warning cues. By providing this continuous representation of thruster state, we hope to enhance situation awareness within the training session and thereby instill 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, tether drag, etc. that will carry over to the real-world task.

This project involves three components: (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; and (3) experimentation to evaluate the proposed display techniques.

An appropriate testbed would include equipment for generating a number of audio source channels having independently controllable waveform parameters (amplitude, frequency, timbre, etc.); some means of spatializing the outputs of the source channels (using filtering, delays, amplitude control, etc.); and any necessary ancillary equipment, such as loudspeakers, headphones, amplifiers, and head trackers, to create the desired acoustic field from the electrical representation of these stimulus signals.

The display formats that we are initially considering 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 distance from the pilot's head. The angular position of the image would correspond to the ROV direction of motion, and the intensity and timbre of the thruster sound would provide some 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 as well. 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 begin to do really precise spatialization), it might be more efficient to simply include a visual indication of the ROV velocity vector on the video display if this were the major consideration.

Since the inception of this research during the latter portion of the time period covered in this report, we have replaced the previous (non-functional) dual computer (PC plus Macintosh) based sound equipment with a more streamlined system consisting of a single PC running three SoundBlaster cards yielding a total of six channels, a Convolvotron, and an Ascension head tracker. The system is configured to operate under Windows NT, and software has been written to implement the two alternative display approaches. 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 condition, we are using the multiple outputs of the SoundBlaster cards mixed (with simple filtering and amplitude tweaking) into the left and right audio channels to create the test stimulus. We are now attempting to determine the best procedures for comparing the alternative display approaches. Once this issue is settled, we will conduct a series of comparative experiments within the dedicated audio testbed. We will then begin to incorporate the newly-developed audio display into the complete training system for more complete testing in realistic training scenarios.

1.9.5 Haptic Force Feedback

The 1 degree-of-freedom (dof) force feedback training experiments that were initiated by Walter Aviles and described in the previous annual report are now being continued. In these experiments, a human test subject controls a simulated ROV represented by a spot of light on a computer monitor. The subject controls the simulated ROV, which is constrained to motion in only one direction, by applying simulated thrust via a joystick. The joystick is a PHANToM manipulandum that is capable of providing various types of force feedback as specified by the experimenter. In the normal or control condition of the experiment, the PHANToM acts like a common spring-return joystick where the force on the subject's hand is proportional to the thrust. The goal of the experiments is to see if some type of additional force feedback during training may enhance the subject's performance, and to determine whether training transfer occurs when the additional force feedback is removed.

In the training strategy now being employed, the normal force/position relationship of the joystick is maintained in a small region about some optimal control position. When the subject's control input deviates too far from the optimum point, however, the spring constant of the joystick increases abruptly, giving the subject a haptic sensation that he/she is out of bounds and physically pushing his/her hand back toward the desired region. In this sense, the proposed strategy is like 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. Note that the optimum position varies with time depending on the task and the performance metric that is used to define "optimum."

In the experiment now being initiated, performance with a normal spring-return joystick is compared following training under three different conditions: (1) the proposed force feedback strategy, (2) a similar training strategy where audio feedback is substituted for the additional force feedback, and (3) a control condition where the normal spring-return joystick is used for training (as well as for the transfer condition).

Two major subtasks are being investigated in connection with this experiment: (1) design of an optimal controller for the task which may be used as a reference in training the subject and for analyzing the results, and (2) establishment of a performance metric suitable for analyzing test results. With regards to the first task, a PID controller has been designed that approaches optimal performance with respect to minimizing position error. With regards to the second task, we are attempting to use mean squared position error to structure the experiments and data analysis in a manner that prevents the effects of the errors that are unlikely to be related to training condition from swamping out the effects of the errors of interest.

We expect to have preliminary results of these experiments within the next two months.

1.10 Training Spatial Knowledge Acquisition Using Virtual Environments

Sponsor

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

Project Staff

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

1.10.1 Overview

The overall objective of this work, which is being conducted in collaboration with Rudy Darken at the Naval Postgraduate School, is to apply state-of-the-art virtual reality simulation technologies to the investigation of issues in spatial knowledge acquisition and training. Virtual environment training provides a number of advantages over traditional training methods and it is expected that such training will translate into measurable real-world gains in performance. The first set of studies are intended to explore the extent to which virtual environments are an effective means of conveying spatial knowledge of a specific real space. The second set, which will not be initiated until later on in the program, will be concerned with the training of spatial knowledge and navigational skills in general and with the ability to transform spatial information presented in one form to information presented in other forms. Throughout all our work, we will explore the role of various environmental features and stimulus characteristics in navigation and wayfinding so that we determine which components of virtual-environment fidelity can be degraded without causing substantial reductions in training transfer. In this manner, we hope to develop systems that are optimally cost effective. To the extent possible, we will also attempt to shape our research to the needs of special communities (such as special operations groups) for whom detailed accurate spatial knowledge of spaces that are not accessible for training is of great practical importance.

Because the work involves both basic and applied research and is also highly interdisciplinary involving involves cognitive science, perceptual psychology, virtual-environment design, real-world training methods, etc., it requires close collaboration among a wide variety of individuals. General background on the use of virtual environments for spatial training can be found in our journal.[59] Further background on this specific project can be found in *Progress Report* No. 139.[60]

During the past year, this work has focused on performing a preliminary experiment comparing various training methods for acquisition of configurational spatial knowledge about the seventh floor of Building 36 at MIT. Configurational knowledge was measured by the ability of subjects to estimate the bearing and range of various landmarks from various stations in the space, where the landmarks and stations were chosen such that the landmarks could not be seen from the stations. Before testing, subjects were allowed to freely explore the space in order to gain familiarization with it. In one case, this exploration took place in the real world. In three other cases, the exploration took place in virtual, computer-generated environments that were constructed to represent the real world. In one such case, an immersive VE was used; in another, a non-immersive VE was used; and in the third, a small scale model was used. Different sets of subjects were used for the different training experiences.

The results of this experiment showed no significant differences among the four training methods: not only were all the VE methods roughly equivalent to each other, but they were all as good as training in the real world. Also of interest in the results of this experiment

59 PRESENCE 7(2): (1998).

60 N.I. Durlach et al., "Training Spatial Knowledge Acquisition Using Virtual Environments," *MIT RLE Progress Report* 139: 374-80 (1996).

was the fact that the variance due to intersubtask differences (i.e., differences in the choice of landmark and station) was roughly comparable to the variance due to intersubject differences.

Further details on this experiment can be found in Koh.[61] A summary of this material is now being prepared for publication.

Results on the development of our robotic room scanner and on the modeling of more complex spaces[62] will be described next year.

1.10.2 Thesis

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

1.11 Further Research on Superauditory Localization for Improved Human-Machine Interfaces

Sponsor

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

Project Staff

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

1.11.1 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. To the extent that the research is successful, it will both advance our understanding of auditory localization and adaptation, and improve our ability to design human-machine interfaces that provide a high level of performance.

During the past year, progress has been made in three areas. First, we have completed our study of adaptation to the azimuthal transformation discussed in *Progress Report* No. 139.[63] Second, we have completed our study of auditory localization in the near field.[64] Finally, we have initiated a new series of experiments on adaptation to an enlarged head in which the simulation of an enlarged head is achieved by frequency scaling the head-related transfer functions.

1.11.2 Publications

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

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

Shinn-Cunningham, B.G. "Adapting to Supernormal Auditory Localization Cues: a Decision-theory Model." *Percept. Psychophys.* Under revision.

Thesis

Brungart, D.S. *Near-Field Auditory Localization.* Ph.D. diss., Department of Electrical Engineering and Computer Science, MIT, 1998.

1.12 Role of Skin Biomechanics in Mechanoreceptor Response

Sponsor

National Institutes of Health
Grant RO1 NS33778

61 G. Koh, *Training Spatial Knowledge Acquisition Using Virtual Environments*, M.Eng. thesis, Department of Electrical Engineering and Computer Science, MIT, 1997.

62 N.I. Durlach et al., "Training Spatial Knowledge Acquisition Using Virtual Environments," *MIT RLE Progress Report* 139: 374-80 (1996).

63 N.I. Durlach et al., "Research on Super-Auditory Localization for Improved Human-Machine Interfaces," *MIT RLE Progress Report* 139: 381-87 (1996); B.G. Shinn-Cunningham, N.I. Durlach, and R.M. Held, "Adapting to Supernormal Auditory Localization Cues I: Bias and Resolution," *J. Acoust. Soc. Am.*, forthcoming; B.G. Shinn-Cunningham, N.I. Durlach, and R.M. Held, "Adapting to Supernormal Auditory Localization Cues II: Changes in Mean Response," *J. Acoust. Soc. Am.*, forthcoming; B.G. Shinn-Cunningham, "Adapting to Supernormal Auditory Localization Cues: a Decision-theory Model," *Percept. Psychophys*, under revision.

64 D.S. Brungart, *Near-Field Auditory Localization*, Ph.D. diss., Department of Electrical Engineering and Computer Science, MIT, 1998.

Project Staff

Dr. Mandayam A. Srinivasan, Balasundar I. Raju, Suvranu De, Amanda S. Birch, Kimberly J. Voss, Joshua P. Cysyk, Dr. Robert H. LaMotte[65]

1.12.1 Overview

Mechanics of the skin and subcutaneous tissues is as central to the sense of touch as the optics of the eye is to vision and the 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 toward 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 three-dimensional (3D) 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) and computer-controlled stimulators; and (4) to collaborate with Professor Robert LaMotte of Yale University School of Medicine in obtaining and analyzing peripheral neural response data for a variety of tactile stimuli.

We have achieved most of our goals except for (1) dynamic stimulation of viscoelastic fingertip models to identify the role of SAI and RA mechanoreceptors in tactile coding, which we will be completing during the remaining project period, and (2) incorporating papillary ridges in the 3D models for which the MRI data did not have as much resolution and contrast as we had hoped. Instead, we 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 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.

1.12.2 Biomechanics

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

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 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 vol-

65 Professor, Yale University, New Haven, Connecticut.

ume changes during the ramp phase increase linearly with indentor displacement and are independent of velocity; during saw tooth 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, when implemented, 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 (3D) 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 different flat plates, one each 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 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 toward 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 *In Vivo* Human Fingertip 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 (3D) models of human and monkey fingertips with realistic external geometry and multi-layered 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. We achieved high resolution and contrast which enabled us to discriminate tissue layers. An MRI 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.

1.12.3 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 Professor LaMotte of Yale University School of Medicine, we have recorded the responses of SAIs and RAs to a variety of 2D and 3D shapes stroked across the monkey fingerpad. One set of experiments involved 2D "wavy surfaces", i.e., surfaces composed of smooth, alternating convex and concave surfaces of differing radii of curvature. The second set of experiments employed 3D 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 spatial population response (SPR), (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. 3D 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 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 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. 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., piece of fruit), 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 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.

1.12.4 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 can 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 2D and 3D finite element models of the primate fingertip. We summarize below the development of the 3D model, and the biomechanical and neurophysiological results obtained from it.

Development of 3D 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 one 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 2D sections were imported into a solid modeler software (PATRAN) and a 3D 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 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 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 SAI nerve fibers through a simple scaling-threshold model and are thus possible *relevant stimuli* for SAI 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. The relevant stimulus measures, such as the strain energy density, are nonlinear functions of the cartesian stress-strain components. Because of the nonlinearity, a simple inverse transformation cannot be applied. A signal estimation technique using the univariate method used in nonlinear 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 section above 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 3D linear viscoelastic continuum. These parameters were then incorporated into a 2D plane strain and a 3D 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 the next grant period, we propose to investigate stroking by textures and indentation by concave and soft objects. In particular, we would like to explore the role of RAs in tactile coding of dynamic objects.

1.12.5 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 classic example of such nonlinear response is the so-called "tuning curves" which are 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 the description of 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.

1.12.6 Device Design and Construction

Ultrasound Backscatter Microscope for *In Vivo* Imaging of Human Fingertip

One conclusion from 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 axial and lateral resolutions were experimentally determined to be about 20 microns and 150 microns respectively.

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 2D 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 fingertip 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 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. We are now planning to investigate the deformations of the skin in the neighborhood of these two mechanoreceptors to various shaped stimuli we have used before (line loads, rectangular bars, cylinders, etc.).

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.

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 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. Currently, data is being collected on the force response of the human fingerpad using sinusoidal inputs in the linear range with a point stimulus. This data will be used to find a linear, visco-elastic model of the human fingerpad and the limits of its validity.

1.12.7 Publications

Journal Articles

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

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. Neurophysiol.* 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. Neurophysiol.* Forthcoming.

Chapter in a Book

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

Thesis

Voss, K.J. *Investigation of the Internal Geometry and Mechanics of the Human Fingertip, in vivo, using Magnetic Resonance Imaging.* M.S. thesis, Department of Mechanical Engineering, MIT, June 1997.

1.13 Force-Reflecting Soft-Tissue Models for Simulating Laparoscopic Surgical Procedures in Virtual Reality Systems

Sponsor

Massachusetts General Hospital, Center for Innovative Minimally Invasive Therapy Research Fellowship Grant

Project Staff

Dr. Cagatay Basdogan, Dr. Mandayam A. Srinivasan, Dr. Stephen D. Small,[66] Dr. Steven L. Dawson[67], Chi-Hao Ho

1.13.1 Overview

Research in the area of computer assisted surgery and surgical simulation has mainly focused on developing 3D geometrical models of the human body from 2D 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 compli-

66 Massachusetts General Hospital, Boston, Massachusetts.

67 Ibid.

ance characteristics of tissues, and the estimation of appropriate reaction forces to convey to the user a feeling of touch and force sensations.

The hardware components of the set-up include an IBM-compatible PC (200 Hz, PentiumPro) with a high-end 3D graphics accelerator, a force-feedback device (PHANToM from SensAble Technologies Inc.) to simulate haptic sensations, and stereo glasses for 3D visualization of virtual objects. 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 stereoscopically 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 multi-threading 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.

1.13.2 Publication

Basdogan, C., C. Ho, M.A. Srinivasan, S.D. Small, S.L. Dawson. "Force Interactions in Laparoscopic Simulations: Haptic Rendering of Soft Tissues." *Proceedings of the Medicine Meets Virtual Reality (MMVR'98) VI Conference*, San Diego, California, January 19-22, 1998, pp. 385-91.