

particular multimodal interface aims to decrease the user's manual fatigue and increase input efficiency.

This type of hybrid multimodal interface potentially can perform more reliably than a pure passive-tracking system, because the active input mode is available to clarify ambiguous user intentions. Early information from the passive mode also can provide predictive power that enhances system robustness and delivers usability advantages, compared with an active mode alone. In the future, this kind of blended multimodal interface may provide the user with greater transparency, better control, and an improved usability experience, while also supporting broader application functionality.

Finally, as new types of multimodal systems proliferate in the future, they also are increasingly likely to include more than two input modes. This trend already has been initiated within the field of biometrics research, which has combined multiple behavioral modes (e.g., voice and handwriting recognition) with physiological ones (e.g., fingerprints, retinal scans) using sensor fusion technologies (Pankanti, Bolle, & Jain, 2000). The driving goal behind this trend to add modes has been improvement of the reliability of person identification and verification tasks within a wider range of realistic usage conditions. As research progresses in the demanding area of biometric security applications, future work will be needed to transfer the more promising new techniques and architectures to other types of interactive multimodal systems.

5.3. Multimodal Language and Dialogue Processing

To provide a scientific foundation for developing multimodal speech and gesture dialogue systems, a general theory of conversational interaction will be needed, as will relatively media-independent representations of intent, semantic content, and context. The generally accepted media-independent level for formulating the basic principles governing dialogue interaction is *communicative intent*, which often is represented in terms of speech acts (Searle, 1969). The same underlying speech act representations are used in dialogue systems for recognition and production, using techniques from planning (Appelt, 1985; Cohen & Perrault, 1979), plan recognition (Allen & Perrault, 1980; Carberry, 1990), formal logic and automated reasoning (Cohen & Levesque, 1990; Perrault, 1990; Sadek, 1991), and statistical language modeling (Reithinger & Klesen, 1997). Although general purpose representations and theories have been available for a number of years, only recently has a real-time, speech-only dialogue system been developed that operates at the level of the speech act (Bretier & Sadek, 1997) and for which the basic representations and processing techniques are common between input and output. Regarding multimodal output, the same types of speech act representations and planning processes have been developed to coordinate the presentation of text and graphics (Feiner &

McKeown, 1991; Wahlster, Andre, Finkler, Profitlich, & Rist, 1993). Future research will need to address the development of complete *multimodal* dialogue systems that incorporate a level of intent representation suitable for nonspeech modalities such as gesture. Furthermore, there needs to be agreement among researchers on a speech act vocabulary that is sufficient for fostering community-wide comparisons and incremental progress.

Regarding media-independent representations of the content of these speech acts, there has been de facto agreement on feature-frame structures, coupled with rules and temporal constraints, to represent multimodal input. Likewise, similar structures and procedures have been developed for planning and coordinating multimodal output (Feiner & McKeown, 1991). Future research will need to synthesize these approaches, employing common representations of temporal information (Allen, 1984). Furthermore, research is needed to develop general media-independent theories of information presentation, both cognitive and formal, that can be realized directly as architectural principles and functioning software. For example, unified theories will be critical in guiding animated character displays that involve nonverbal gestures and facial expressions as communicative output (Andre, Rist, & Muller, 1999; Cassell & Stone, 1999; Cassell, Sullivan, Prevost, & Churchill, 2000; Lester, Voerman, Towns, & Callaway, 1999).

The glue between input and output is the representation and use of context. Unlike GUIs that are designed to be context independent, systems based on natural language have long been designed to represent and use context, for example, to interpret pronouns. More generally, context will need to be derived from the user's prior *multimodal* interactions, and the system's multimedia presentations. The latter includes the visual context of what is on the screen and, for future systems that employ camera input, the user's physical context. Unfortunately, most multimedia systems simply present information without representing its context. Future systems that *synthesize* multimedia (e.g., Feiner & McKeown, 1991; Roth, Mattis, & Mesnard, 1991) will be needed to capture visual context adequately. Precisely what to save from prior multimodal interaction, and how to deploy these contextual representations to best ensure correct multimodal interpretations, is a crucial subject of future research.

Finally, to arrive at multimodal interpretations, future multimodal dialogue systems should be developed within a statistical framework (Horvitz, 1999) that permits probabilistic reasoning about the task, the context, and typical user intentions, together with statistical language modeling. Given this framework, a complete multimodal dialogue system perhaps could be trained on an annotated corpus of interactions that provides the basis for correlating speech, gesture, natural language, context (linguistic and visual), task scenario, and user intent. To accomplish this, large-scale multimodal corpora need to be de-

veloped and annotated. In addition, statistical interpretation algorithms, such as MTC, will need to be extended and tested on these new corpora.

5.4. Error-Handling Techniques

Fragile error handling currently remains the number one interface problem for recognition-based technologies like speech and pen (Karat et al., 1999; Oviatt, 2000b; Rhyne & Wolf, 1993; Roe & Wilpon, 1994). However, as discussed in Section 2, multimodal interface designs that combine two input modes tend to have superior error-handling characteristics (Oviatt, 1999a, 2000b; Oviatt et al., 1999; Oviatt & vanGent, 1996; Rudnicky & Hauptmann, 1992; Suhm, 1998; Suhm, Myers, & Waibel, 1996). Future research needs to continue to investigate strategies for developing graceful error handling in multimodal systems. In particular, research should explore issues such as (a) methods for designing multimodal architectures that support higher levels of mutual disambiguation between signals, (b) the impact of incorporating a third input mode on error avoidance and system stability, (c) the impact of new language and dialogue-processing techniques on error avoidance and resolution, and (d) adaptive architectures that reduce errors and stabilize system performance in noisy mobile environments and other challenging contexts. Eventually, improved error avoidance and correction techniques also need to become available to application developers in the form of toolkits (Vo, 1998).

5.5. Adaptive Multimodal Architectures

The future research agenda for developing adaptive multimodal speech and gesture architectures subsumes the problems of what and when to adapt, as well as how to adapt multimodal systems so that their robustness can be enhanced. Adaptive multimodal architectures will increase the ease with which users can interact with a system by continually adjusting to a user and his or her surroundings. Two primary candidates for system adaptation are *user-centered* and *environmental* parameters.

With respect to user-centered parameters, previous empirical work on multimodal pen-voice integration has revealed two main types of user—ones who habitually deliver speech and pen signals in an overlapped or simultaneous manner, and others who synchronize signals sequentially with pen input preceding speech by up to 4 sec (Oviatt, 1999b). Any given user's habitual integration pattern is apparent at the beginning of the user's system interaction. As a result, future multimodal systems that are capable of distinguishing these two types of user, and adjusting the system's temporal thresholds for integrating signals accordingly, potentially could achieve greater recognition accuracy and interactive speed.

With respect to environmental parameters for adapting multimodal pen-voice systems, background noise and speech signal-to-noise ratio (SNR) are two widely used audio dimensions that have an impact on system recognition rates. To design robust mobile pen-voice systems for field use, future research will need to experiment with adaptive processing that tracks the background noise level and dynamically adjusts the system's weightings for speech and gesture input from the less to more reliable input mode (Rogozan & Deleglise, 1998). In addition, future research needs to establish an analogue to SNR for calibrating noisy pen input and its impact on gesture recognition rates, which could be especially consequential during movement, fatigue, or device-induced jitter. In general, the quality of speech and pen input signals can be estimated by relative recognition uncertainty, which some system recognizers produce along with posterior probabilities. The MTC approach outlined in Section 3.4 estimates signal-level uncertainty and potentially could be used to provide a base architecture for developing successful adaptive processing.

Most existing adaptive multimodal architectures were developed for audio-visual recognition of speech and lip movements. These architectures have relied on two general types of models, ones that use *parameter-based adaptive weights*, and others based on *black-box adaptivity*. For models based on parameter-based weightings, the system calibrates an individual modality's weighting (e.g., speech) as directly proportional to the magnitude of the relevant environmental parameter (e.g., strength of SNR [Meier, Hurst, & Duchnowski, 1996] or probability dispersion weights [Adjoudani & Benoit, 1996]). That is, if SNR is a low value, then the reliability of the speech signal interpretation would be regarded as weak and this mode would be relatively underweighted compared with gesture input. In addition, different types of multimodal commands can be assigned different weights, just as modalities can (Wu et al., 1999). For black-box models (e.g., NNs), the relation between variables and modality weights is built through supervised learning (Sejnowski, Yuhua, Goldstein, & Jenkins, 1990; Waibel, Vo, Duchnowski, & Manke, 1995). This learned relation is completely dependent on training data and may be highly nonlinear. A major area in need of further investigation is the application of these adaptive processing methods specifically to pen-voice multimodal input.

5.6. Multidevice, Multiuser Ubiquitous Computing

Most of today's multimodal applications have been built around a single device or closely coupled complex of devices. As we move toward an era of ubiquitous computing, people will find themselves in environments in which they are not dealing with a "computer" per se. Rather, they will be moving among a variety of complementary devices, each appropriate for a particular task. For example, in an "interactive room" (Coen, 1998; Pentland, 1996)

there may be display devices of all sizes (e.g., handheld to wall-screen) and orientations (e.g., wall, table). They could be controlled with pointing devices that are direct or indirect (e.g., pen vs. trackball), in contact or distant (e.g., pen vs. laser pointer), or specialized for different kinds of manipulation (e.g., gloves). Cameras and microphones will create a *perceptual space* in which what people say and do is gathered as input signals to be interpreted within the environment's overall computing context.

In future interactive environments such as this, multimodal interaction styles will predominate. One important future research goal will be to understand the nature of multimodal interaction within such a computing context. Another will be to create scalable and reusable architectures that cope effectively with the diversity of these input modes and signals—each with its unique characteristics involving response latency, recognition certainty, adaptability to the user, and so forth. This will include developing standard ways to indicate multiple alternatives for an input signal, and easy configuration–reconfiguration of components for interpreting input based on multiple data sources. In comparison with this vision, today's multimodal speech and gesture systems (e.g., outlined in Section 4) eventually will be viewed as just the first step in a longer term evolution away from systems based on specialized code and devices.

As multimodal interaction transforms into larger physical spaces and ubiquitous forms, multiple users often will be interacting together—pointing to a shared display surface, gesturing simultaneously, and talking in a natural style that includes interruptions, disfluencies, overlapped speech, and so forth. To date, the research in computer-supported collaborative work involving issues like turn taking and dialogue control has dealt primarily with individual communication modes, usually textual in nature. Such work will need to be extended to handle natural *multimodal* interactions, including the technological issues, but also understanding how groups of people will communicate and coordinate when using combined speech and gesture to accomplish daily computing tasks. If multimodal spaces of this kind are to become commonly accepted, they also will have to minimize the amount or size of gear that a user wears or carries. This will require greater dependence on ambient environmental devices, such as wide-field cameras and directional microphone arrays (Flanagan, Berkeley, Elko, West, & Sondhi, 1991). It also will require the development of methods for tracking and recognizing an individual user's contributions to the overall signal environment. Among other things, this will necessitate basic changes in the way that operating systems deal with input, because today's systems generally implicitly assume that a particular device provides input from a single individual at a given moment. In sum, the prospect of future multidevice, multiuser ubiquitous computing presents a wide array of long-term research challenges.

5.7. Multimodal Research Infrastructure

For the fledgling multimodal research community to develop high-performance multimodal systems and eventually commercialize them, considerable research will be needed to develop appropriate infrastructure, including (a) semiautomatic simulation methods for empirical data collection and prototyping new systems, (b) automated tools for collecting and analyzing multimodal corpora during realistic usage contexts, (c) novel metrics for evaluating multimodal systems, and (d) automated corpus collection and analysis tools for iterating new multimodal systems to steadily expand and improve their performance. Before multimodal systems can proliferate, the community also will need (e) software tools that support the rapid creation of next-generation multimodal applications.

Tools for Multimodal System Design, Iteration, and Corpus Collection

Multimodal systems still are very new and hard to build. Among other things, advancing this research area will require proactive research and situated data collection to achieve high-quality multimodal system design. More specifically, software support is needed for developing semiautomatic simulation techniques, which are the preferred method for designing multimodal prototypes for systems still in the planning stages (Cheyer, Julia, & Martin, 1998; Oviatt, 1996; Oviatt et al., 1992). In a simulation environment, the user believes that she is interacting with a fully functional system, while a trained programmer assistant actually provides simulated system responses from a remote location. Simulation software is designed to support a subject-paced rapid interaction with the simulated system. It also is designed to be rapidly adaptable, so that alternative interface designs and system features can be investigated fully and easily. Simulation-based research can help designers to evaluate critical performance tradeoffs and make decisions about a system's design, which is necessary for creating usable multimodal systems that have valuable functionality.

In the future, simulation tools will need to be developed that permit researchers to explore new input modes and interface designs appropriate for a wide spectrum of different multimodal systems. There also is a special need for longitudinal user studies, especially on multimodal interfaces that incorporate novel media—ones with which users have limited familiarity. In addition, because there is strong interest in designing multimodal systems for use in natural field environments and while users are mobile, simulation and other data-collection tools are especially needed for supporting situated data collection in such contexts (e.g., in-vehicle and emergency medical applications; see Holzman, 1999). New tools are just beginning to be developed for this purpose (Oviatt & Pothering, 1998).

In the area of spoken-language processing, the availability of significant corpora of transcribed and annotated training data has been a critical resource enabling the remarkably rapid progress in spoken-language understanding during the past decade (Cole et al., 1995). Unfortunately, the unavailability of common multimodal corpora currently is a problem that has been impeding progress on the development of new methods for multimodal language processing (i.e., including linguistic and statistical techniques), as well as techniques for multimodal signal fusion and adaptive processing. More widespread availability of multimodal corpora also is critically needed to develop appropriate methods for evaluating and iterating the overall performance of new multimodal architectures. The collection and analysis of large multimodal corpora could be expedited greatly by the development of automated multimodal data loggers. A few such multimodal data loggers are beginning to appear for recording, searching, analyzing, and generating automatic summaries about the signal-level and linguistic features of users' multimodal utterances, as well as system recognition results on multimodal input (Clow & Oviatt, 1998).¹⁵

Tools for Multimodal Software Development

Multimodal systems are more complex than unimodal ones, making system design and implementation more difficult. Moreover, recognizers for various modalities can be difficult to acquire and to integrate because toolkit-level support is limited at best, and integration of recognizers in new applications requires significant expertise. There presently are no software tools to assist application designers with building multimodal applications. Only recently have programming toolkits become available for recognizing natural unimodal input such as speech (Huang et al., 1995; Kempf, 1994; Sun Microsystems, 1998). These new single-mode toolkits usually are intended for use by recognition technology specialists or advanced programmers. However, an example of an exception to this is the CSLURp interactive builder for speech applications (Sutton & Cole, 1997), which aims to provide tools for nonspecialists, although it is limited to speech-only transaction interfaces.

In the future, developers will need prototyping tools for designing multimodal interfaces that combine two or more natural modes, possibly in conjunction with conventional direct manipulation. These tools ideally should be multimodal themselves, as well as incorporating programming by demonstration (Cypher, 1993) so that designers with limited programming experience can prototype multimodal designs rapidly. For example, using programming by demonstration techniques, a designer potentially could specify speech and

15. Also see MITRE's logger at <http://www.mitre.org/research/logger/release/1.0/html/logger.html>

gesture input operations for a "Wizard of Oz" simulation, which could enable the simulation system to infer the user's multimodal behavior during testing, thereby reducing the human wizard's workload significantly.

Prototyping tools also should support the informal, iterative techniques that designers currently use in the early stages of design—from sketching visual interface elements using a system like SILK (Landay, 1996; Landay & Myers, 1995, 1996), to storyboarding, to high-fidelity "Wizard of Oz" simulation techniques originally developed for spoken-language (Kelley, 1984) and more recently for multimodal systems (Cheyer et al., 1998; Oviatt et al., 1992). These multimodal prototyping tools also should give designers the flexibility to fill in design details as choices are made (Wagner, 1990) and test the designs interactively. For example, interactive tools should permit the designer to demonstrate "Wizard of Oz" sequences that simulate the intended behavior of the recognizer, as the SUEDE system is intended to do for spoken language (Klemmer, Sinha, Chen, Landay, & Aboobaker, 2000). Together, software tools that support prototyping techniques such as interactive sketching, programming by demonstration, and high-fidelity simulations can be used to design multimodal systems far more rapidly, which will be necessary for this new generation of interfaces to proliferate.

6. CONCLUSIONS

Multimodal systems that process users' speech and pen-based gestural input have become a vital and expanding field, especially within the past 5 to 8 years, with demonstrated advances in a growing number of research and application areas. Among the benefits of multimodal interface design are general facilitation of the ease, flexibility, and power of computing, support for more challenging applications and forms of computing than in the past, the expanded use of computing while mobile and in natural field settings, and a potentially major increase in the accessibility of computers to a wider and more diverse range of users. In particular, because multimodal systems support relatively natural interaction without special training, they will make computing and information services available to field workers, business and service people, students, travelers and other mobile users, children, the elderly, permanently and temporarily disabled users, and computer-illiterate and casual users.

In this article, we summarized the prevailing and emerging architectural approaches available for interpreting dual input signals in a robust manner, including early and late semantic fusion approaches, as well as a new hybrid symbolic-statistical architecture for processing pen-voice input. We also described a diverse collection of state-of-the-art multimodal systems that are capable of processing users' spoken and gestural input—ranging from map-based and VR systems for engaging in simulations and training to field medic sys-

tems for mobile use in noisy environments, to Web-based transactions and standard text-editing applications that will reshape daily computing tasks. We indicated that many key research challenges remain to be addressed before successful multimodal systems can be realized fully. Among these challenges are the development of cognitive theories of multimodal interaction, as well as more effective natural language, dialogue processing, semantic integration, and error-handling techniques. Although the rapid maturation of SLT has contributed directly to recent advances in multimodal systems, further strides will be needed in other component technologies and hardware before existing multimodal systems can be diversified and optimized fully. In addition, new and more sophisticated architectures will be needed for handling media-independent representations; designing more robust and adaptive multimodal systems; and supporting multidevice, multiuser use. Before this new class of systems can proliferate, multimodal toolkits also will be needed to support software development, as will simulation-based methods for corpus collection, analysis, and effective system iteration.

Finally, some of the requirements for advancing innovative multimodal systems are not intellectual ones, but rather social, political, and educational in nature. The development of state-of-the-art multimodal systems of the kind described in this article also requires multidisciplinary expertise in a variety of areas, such as speech and hearing science, perception and graphics, linguistics, psychology, signal processing, pattern recognition, statistics, engineering, and computer science. The multidisciplinary nature of this research makes it unlikely that a single group can conduct meaningful research across the entire spectrum. As a result, collaborative research and "community building" among multimodal researchers will be critically needed to forge the necessary relations among those representing different component technologies and key disciplines. In addition to cross-fertilization of ideas and perspectives among these diverse groups, there also is a critical need for cross-training of students and junior researchers. Like spoken-language systems, multimodal technology does not fit neatly into a traditional academic departmental framework. To make the appropriate educational opportunities and resources available to future students, new academic programs certainly will need to be formulated that encourage and reward researchers who successfully reach across the boundaries of their narrowly defined fields.

NOTES

Background. This article is based partly on discussions held at a CHI'99 Workshop meeting in Pittsburgh, Pennsylvania during May 1999. The workshop, "Designing the User Interface for Pen and Speech Multimedia Applications," was

organized by James A. Larson (Intel Architecture Labs), David Ferro (Unisys NLU group), and Sharon Oviatt (Computer Science Department at Oregon Graduate Institute). In addition to its authors, this article also benefitted from workshop discussions with Adam Cheyer of SRI, Christopher Esposito of Boeing, Jennifer Mankoff of Georgia Tech, Mark Newman of University of California-Berkeley, Pernilla Qvarfordt of Linköping University in Sweden, Dan Venolia of Microsoft, and Timothy Miller of Brown University.

Acknowledgments. We thank Dana Director, Gwen Brake, and Bridget Adams for their expert assistance with graphics and manuscript preparation.

Support. The preparation of this article has been supported in part by Intel Corporation, by National Science Foundation (NSF) Grant IRI-9530666 and NSF Special Extension for Creativity Grant IIS-9530666 to Sharon Oviatt. Additional support was also provided by Defense Advanced Research Projects Agency contracts DABT63-95-C-007 and N66001-99-D-8503, as well as Office of Naval Research Grant N0014-95-1164 to Phil Cohen.

First Author's Present Address. Sharon Oviatt, Center for Human-Computer Communication, Department of Computer Science and Engineering, Oregon Graduate Institute of Science & Technology, 20000 N.W. Walker Road, Beaverton, OR 97006. E-mail: oviatt@cse.ogi.edu.

HCI Editorial Record. First manuscript received August 6, 1999. Revision received May 2000. Accepted by Clayton Lewis. Final manuscript received July 13, 2000. — *Editor*

REFERENCES

- Adjoudani, A., & Benoit, C. (1996). On the integration of auditory and visual parameters in an HMM-based ASR. In D. Stork & M. Hennecke (Eds.), *Speechreading by humans and machines, NATO ASI series, series F: Computer and systems science* (pp. 461-473). Berlin: Springer.
- Allen, J. F. (1984). Towards a general theory of action and time. *Artificial Intelligence*, 23, 123-154.
- Allen, J. F., & Perrault, C. R. (1980). Analyzing intention in dialogues. *Artificial Intelligence*, 15, 143-178.
- Andre, E., Rist, T., & Muller, J. (1999). Employing AI methods to control the behavior of animated interface agents. *Applied Artificial Intelligence*, 13, 415-448.
- Appelt, D. (1985). *Planning English sentences*. Cambridge, England: Cambridge University Press.
- Benoit, C., Martin, J.-C., Pelachaud, C., Schomaker, L., & Suhm, B. (2000). Audio-visual and multimodal speech systems. In D. Gibbon, I. Mertins, & R. Moore (Eds.), *Handbook of multimodal and spoken dialogue systems: Resources, terminology and product evaluation*. Norwell, MA: Kluwer Academic.
- Bers, J., Miller, S., & Makhoul, J. (1998). Designing conversational interfaces with multimodal interaction. *DARPA Workshop on Broadcast News Understanding Systems*, 319-321.
- Bolt, R. A. (1980). Put-that-there: Voice and gesture at the graphics interface. *Computer Graphics*, 14, 262-270.

- Bregler, C., Manke, S., Hild, H., & Waibel, A. (1993). Improving connected letter recognition by lipreading. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, 1, 557-560. Minneapolis: IEEE Press.
- Bretier, P., & Sadek, D. (1997). A rational agent as the kernel of a cooperative spoken dialogue system: Implementing a logical theory of interaction. *Intelligent Agents III: Proceedings of the Third International Workshop on Agent Theories, Architectures and Languages (ATAL-96)*, 189-204. Heidelberg, Germany: Springer-Verlag.
- Calder, J. (1987). Typed unification for natural language processing. In E. Klein & J. van Benthem (Eds.), *Categories, polymorphisms, and unification* (pp. 65-72). Edinburgh, Scotland: Center for Cognitive Science, University of Edinburgh.
- Carberry, S. (1990). *Plan recognition in natural language dialogue*. Cambridge, MA: ACL-MIT Press Series in Natural Language Processing, MIT Press.
- Carpenter, R. (1990). Typed feature structures: Inheritance, (in)equality, and extensionality. *Proceedings of the ITK Workshop: Inheritance in Natural Language Processing*, 9-18. Tilburg, The Netherlands: Institute for Language Technology and Artificial Intelligence, Tilburg University.
- Carpenter, R. (1992). *The logic of typed feature structures*. Cambridge, England: Cambridge University Press.
- Cassell, J., & Stone, M. (1999). Living hand to mouth: Psychological theories about speech and gesture in interactive dialogue systems. *Working notes of the AAAI'99 Fall Symposium Series on Psychological Models of Communication in Collaborative Systems*, 34-42. Menlo Park, CA: AAAI Press.
- Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (Eds.). (2000). *Embodied conversational agents*. Cambridge, MA: MIT Press.
- Cheyner, A., & Julia, L. (1995). Multimodal maps: An agent-based approach. *International Conference on Cooperative Multimodal Communication (CMC'95)*, 103-113. Eindhoven, The Netherlands: Eindhoven University of Technology.
- Cheyner, A., Julia, L., & Martin, J. C. (1998). A unified framework for constructing multimodal experiments and applications. *Conference on Cooperative Multimodal Communication (CMC'98)*, 63-69. Tilburg, The Netherlands: Institute for Language Technology and Artificial Intelligence, Tilburg University.
- Clow, J., & Oviatt, S. L. (1998). STAMP: A suite of tools for analyzing multimodal system processing. *Proceedings of the International Conference on Spoken Language Processing*, 2, 277-280. Sydney, Australia: ASSTA, Inc.
- Codella, C., Jalili, R., Koved, L., Lewis, J., Ling, D., Lipscomb, J., Rabenhorst, D., Wang, C., Norton, A., Sweeney, P., & Turk, C. (1992). Interactive simulation in a multi-person virtual world. *Proceedings of the Conference on Human Factors in Computing Systems (CHI'92)*, 329-334. New York: ACM Press.
- Coen, M. (1998). Design principles for intelligent environments. *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI'98)*, 547-554. Menlo Park, CA: AAAI Press.
- Cohen, P. R., Cheyner, A., Wang, M., & Baeg, S. C. (1997). An open agent architecture. In M. Huhns & M. Singh (Eds.), *Readings in agents* (pp. 197-204). San Francisco: Kaufmann. (Reprinted from *AAAI '94 Spring Symposium Series on Software Agents*, 1-8, 1994, Menlo Park, CA: AAAI Press)
- Cohen, P. R., Dalrymple, M., Moran, D. B., Pereira, F. C. N., Sullivan, J. W., Gargan, R. A., Schlossberg, J. L., & Tyler, S. W. (1998). Synergistic use of direct manipula-

- tion and natural language. In M. Maybury & W. Wahlster (Eds.), *Readings in intelligent user interfaces* (pp. 29–37). San Francisco: Kaufmann. (Reprinted from *Proceedings of the Conference on Human Factors in Computing Systems (CHI'89)*, 227–234, 1989, New York: ACM Press)
- Cohen, P. R., Johnston, M., McGee, D., Oviatt, S. L., Clow, J., & Smith, I. (1998). The efficiency of multimodal interaction: A case study. *Proceedings of the International Conference on Spoken Language Processing*, 2, 249–252. Sydney, Australia: ASSTA, Inc.
- Cohen, P. R., Johnston, M., McGee, D., Oviatt, S., Pittman, J., Smith, I., Chen, L., & Clow, J. (1997). Quickset: Multimodal interaction for distributed applications. *Proceedings of the Fifth ACM International Multimedia Conference*, 31–40. New York: ACM Press.
- Cohen, P. R., & Levesque, H. J. (1990). Rational interaction as the basis for communication. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 221–255). Cambridge, MA: SDF Benchmark Series, MIT Press.
- Cohen, P. R., McGee, D., Oviatt, S., Wu, L., Clow, J., King, R., Julier, S., & Rosenblum, L. (1999). Multimodal interaction for 2D and 3D environments. *IEEE Computer Graphics and Applications*, 19(4), 10–13.
- Cohen, P. R., & Oviatt, S. L. (1995). The role of voice input for human-machine communication. *Proceedings of the National Academy of Sciences*, 92, 9921–9927. Washington, DC: National Academy of Sciences Press.
- Cohen, P. R., & Perrault, C. R. (1979). Elements of a plan-based theory of speech acts. *Cognitive Science*, 3(3), 177–212.
- Cole, R., Hirschman, L., Atlas, L., Beckman, M., Biermann, A., Bush, M., Clements, M., Cohen, P., Garcia, O., Hanson, B., Hermansky, H., Levinson, S., McKeown, K., Morgan, N., Novick, D., Ostendorf, M., Oviatt, S., Price, P., Silverman, H., Spitz, J., Waibel, A., Weinstein, C., Zahorian, S., & Zue, V. (1995). The challenge of spoken language systems: Research directions for the nineties. *IEEE Transactions on Speech and Audio Processing*, 3(1), 1–21.
- Cole, R., Mariani, J., Uszkoreit, H., Varile, G., Zaenen, A., Zue, V., & Zampolli, A. (Eds.). (1997). *Survey of the state of the art in human language technology*. Cambridge, MA: Cambridge University Press.
- Cypher, A. (Ed.). (1993). *Watch what I do: Programming by demonstration*. Cambridge, MA: MIT Press.
- Duncan, L., Holmback, H., Kau, A., Poteet, S., Miller, S., Harrison, P., Powell, J., & Jenkins, T. (1994). *Message processing and data extraction for the Real-Time Information Extraction System (RTIMS) project* (Boeing Technical Report BCSTECH-94-057).
- Duncan, L., Brown, W., Esposito, C., Holmback, H., & Xue, P. (1999). *Enhancing virtual maintenance environments with speech understanding*. Boeing M&CT TechNet.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.
- Ermann, L. D., & Lesser, V. R. (1975). A multi-level organization for problem solving using many diverse sources of knowledge. *Proceedings of the 4th International Joint Conference on Artificial Intelligence*, 483–490. Georgia, USSR: IJCAI Press.
- Feiner, S., MacIntyre, B., Hollerer, T., & Webster, A. (1997). A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. *International Symposium on Wearable Computing (ISWC'97)*, 208–217. Cambridge, MA: CS Press.

- Feiner, S. K., & McKeown, K. R. (1991). COMET: Generating coordinated multimedia explanations. *Proceedings of the Conference on Human Factors in Computing Systems (CHI'91)*, 449-450. New York: ACM Press.
- Fell, H., Delta, H., Peterson, R., Ferrier, L., Mooraj, Z., & Valteau, M. (1994). Using the baby-babble-blanket for infants with motor problems. *Proceedings of the Conference on Assistive Technologies (ASSETS'94)*, 77-84. Marina del Rey, CA. Retrieved from the World Wide Web: <http://www.acm.org/sigcaph/assets/assets98/assets98index.html>
- Flanagan, J., Berkeley, D., Elko, G., West, J., & Sondhi, M. (1991). Auto-directive microphone systems. *Acustica*, 73, 58-71.
- Fukumoto, M., Suenaga, Y., & Mase, K. (1994). Finger-pointer: Pointing interface by image processing. *Computer Graphics*, 18, 633-642.
- Goldschen, A., & Loehr, D. (1999). The Role of the DARPA communicator architecture as a human computer interface for distributed simulations. *Spring Simulation Interoperability Workshop*. Orlando, FL: Simulation Interoperability Standards Organization.
- Gong, Y. (1995). Speech recognition in noisy environments: A survey. *Speech Communication*, 16, 261-291.
- Hauptmann, A. G. (1989). Speech and gestures for graphic image manipulation. *Proceedings of the International Conference on Human-Computer Interaction (CHI'89)*, 1, 241-245. New York: ACM Press.
- Holmback, H., Duncan, L., & Harrison, P. (2000). A word sense checking application for Simplified English. *Proceedings of the Third International Workshop on Controlled Language Applications*, Seattle.
- Holmback, H., Greaves, M., & Bradshaw, J. (1999). A pragmatic principle for agent communication. In J. Bradshaw, O. Etzioni, & J. Mueller (Eds.), *Proceedings of Autonomous Agents '99*. New York: ACM Press.
- Holzman, T. G. (1999). Computer-human interface solutions for emergency medical care. *Interactions*, 6(3), 13-24.
- Horvitz, E. (1999). Principles of mixed-initiative user interfaces. *Proceedings of the Conference on Human Factors in Computing Systems (CHI'99)*, 159-166. New York: ACM Press.
- Huang, X., Acero, A., Alleva, F., Hwang, M.-Y., Jiang, L., & Mahajan, M. (1995). Microsoft Windows highly intelligent speech recognizer: Whisper. *Proceedings of 1995 International Conference on Acoustics, Speech, and Signal Processing*, 1, 93-96. Detroit, MI: IEEE.
- Iverson, P., Bernstein, L., & Auer, E. (1998). Modeling the interaction of phonemic intelligibility and lexical structure in audiovisual word recognition. *Speech Communication*, 26(1&2), 45-63.
- Johnston, M. (1998). Unification-based multimodal parsing. *Proceedings of the International Joint Conference of the Association for Computational Linguistics and the International Committee on Computational Linguistics (COLING-ACL'98)*, 624-630. Montreal, Canada: University of Montreal Press.
- Johnston, M., Cohen, P. R., McGee, D., Oviatt, S. L., Pittman, J. A., & Smith, I. (1997). Unification-based multimodal integration. *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, 281-288. San Francisco: Kaufmann.

- Juang, B.-H., & Katagiri, S. (1992). Discriminative learning for minimum error classification. *IEEE Transactions on Signal Processing*, 40(12), 3043-3054.
- Karat, C.-M., Halverson, C., Horn, D., & Karat, J. (1999). Patterns of entry and correction in large vocabulary continuous speech recognition systems. *Proceedings of the International Conference for Computer-Human Interaction (CHI'99)*, 568-575. New York: ACM Press.
- Karshmer, A. I., & Blattner, M. (Organizers). (1998). *Proceedings of the Third International ACM Proceedings of the Conference on Assistive Technologies (ASSETS'98)*. Marina del Rey, CA. <http://www.acm.org/sigcaph/assets/assets98/assets98index.html>
- Katagiri, S., & McDermott, E. (in press). Discriminative training—recent progress in speech recognition. In C. H. Chen, L. F. Pau, & P. S. P. Wang (Eds.), *Handbook of pattern recognition and computer vision* (2nd ed., No. 3). World Scientific Publishing Company.
- Kay, M. (1979). Functional grammar. *Proceedings of the Fifth Annual Meeting of the Berkeley Linguistics Society*, 142-158. Berkeley, CA: Berkeley Linguistics Society.
- Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Office Information Systems*, 2, 26-41.
- Kempf, J. (1994). *Preliminary handwriting recognition engine application program interface for Solaris 2*. Mountain View, CA: Sun Microsystems.
- Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207-227). The Hague, The Netherlands: Mouton.
- Klemmer, S., Sinha, A., Chen, J., Landay, J., & Aboobaker, A. (2000). SUEDE: A Wizard of Oz prototyping tool for speech user interfaces. *Proceedings of User Interface Software and Technology 2000*, 1-10. San Diego, CA: ACM Press.
- Kumar, S., & Cohen, P. R. (2000). Towards a fault-tolerant multi-agent system architecture. *Fourth International Conference on Autonomous Agents 2000*, 459-466. Barcelona: ACM Press.
- Lai, J., & Vergo, J. (1997). MedSpeak: Report creation with continuous speech recognition. *Proceedings of the Conference on Human Factors in Computing (CHI'97)*, 431-438. New York: ACM Press.
- Landay, J. A. (1996). *Interactive sketching for the early stages of user interface design*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.
- Landay, J. A., & Myers, B. A. (1995). Interactive sketching for the early stages of user interface design. *Proceedings of the Conference on Human Factors in Computing Systems (CHI'95)*, 43-50. New York: ACM Press.
- Landay, J. A., & Myers, B. A. (1996). Sketching storyboards to illustrate interface behavior. *Human Factors in Computing Systems (CHI'96), Conference Companion*, 193-194. New York: ACM Press.
- Larson, J. A., Oviatt, S. L., & Ferro, D. (1999). Designing the user interface for pen and speech applications. *CHI'99 Workshop, Conference on Human Factors in Computing Systems (CHI'99)*. New York: ACM Press.
- Lester, J., Voerman, J., Towns, S., & Callaway, C. (1999). Deictic believability: Coordinated gesture, locomotion, and speech in lifelike pedagogical agents. *Applied Artificial Intelligence*, 13, 383-414.

- Levelt, W., Richardson, G., & Heu, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language*, 24, 133-164.
- Martin, A., Fiscus, J., Fisher, B., Pallett, D., & Przybicki, M. (1997). System descriptions and performance summary. *Proceedings of the Conversational Speech Recognition Workshop/DARPA Hub-5E Evaluation*. San Francisco: Kaufmann.
- Martin, D. L., Cheyer, A. J., & Moran, D. B. (1999). The open agent architecture: A framework for building distributed software systems. *Applied Artificial Intelligence*, 13, 91-128.
- McGee, D., & Cohen, P. R. (2001). Creating tangible interfaces by augmenting physical objects with multimodal language. *Proceedings of the International Conference on Intelligent User Interfaces*, 113-119. Santa Fe, NM: ACM Press.
- McGee, D., Cohen, P. R., & Oviatt, S. L. (1998). Confirmation in multimodal systems. *Proceedings of the International Joint Conference of the Association for Computational Linguistics and the International Committee on Computational Linguistics (COLING-ACL'98)*, 823-829. Montreal, Canada: University of Montreal Press.
- McGee, D. R., Cohen, P. R., & Wu, L. (2000). Something from nothing: Augmenting a paper-based work practice with multimodal interaction. *Proceedings of the Designing Augmented Reality Environments Conference 2000*, 71-80. Copenhagen: ACM Press.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- Meier, U., Hurst, W., & Duchnowski, P. (1996). Adaptive bimodal sensor fusion for automatic speechreading. *Proceedings of International Conference on Acoustic, Speech and Signal Processing*, 2, 833-836. IEEE Press.
- Minsky, M. (1975). A framework for representing knowledge. In P. Winston (Ed.), *The psychology of computer vision* (pp. 211-277). New York: McGraw-Hill.
- Neal, J. G., & Shapiro, S. C. (1991). Intelligent multimedia interface technology. In J. Sullivan & S. Tyler (Eds.), *Intelligent user interfaces* (pp. 11-43). New York: ACM Press.
- Nicolino, T. (1994). A natural language processing based situation display. *Proceedings of the 1994 Symposium on Command and Control Research and Decision Aids*, 575-580. Monterey, CA: Naval Postgraduate School.
- Oviatt, S. L. (1992). Pen/voice: Complementary multimodal communication. *Proceedings of Speech Tech'92*, New York.
- Oviatt, S. L. (1996). User-centered design of spoken language and multimodal interfaces. In M. Maybury & W. Wahlster (Eds.), *Readings on intelligent user interfaces* (pp. 620-629). San Francisco: Kaufmann. (Reprinted from *IEEE Multimedia*, 3(4), 26-35, 1996)
- Oviatt, S. L. (1997). Multimodal interactive maps: Designing for human performance. *Human-Computer Interaction*, 12, 93-129.
- Oviatt, S. L. (1999a). Mutual disambiguation of recognition errors in a multimodal architecture. *Proceedings of the Conference on Human Factors in Computing Systems (CHI'99)*, 576-583. New York: ACM Press.
- Oviatt, S. L. (1999b). Ten myths of multimodal interaction. *Communications of the ACM*, 42(11), 74-81.

- Oviatt, S. L. (2000a). Multimodal system processing in mobile environments. *Proceedings of the 13th annual ACM Symposium on User Interface Software and Technology (UIST 2000)*, 21-30. New York: ACM Press.
- Oviatt, S. L. (2000b). Taming recognition errors with a multimodal interface. *Communications of the ACM*, 43(9), 45-51.
- Oviatt, S. L., Bernard, J., & Levow, G. (1999). Linguistic adaptation during error resolution with spoken and multimodal systems. *Language and Speech*, 41(3&4), 415-438.
- Oviatt, S. L., & Cohen, P. R. (1991). Discourse structure and performance efficiency in interactive and noninteractive spoken modalities. *Computer Speech and Language*, 5(4), 297-326.
- Oviatt, S. L., & Cohen, P. R. (2000). Multimodal systems that process what comes naturally. *Communications of the ACM*, 43(3), 45-53.
- Oviatt, S. L., Cohen, P. R., Fong, M. W., & Frank, M. P. (1992). A rapid semi-automatic simulation technique for investigating interactive speech and handwriting. *Proceedings of the International Conference on Spoken Language Processing*, 2, 1351-1354. Alberta, Canada: University of Alberta.
- Oviatt, S. L., Cohen, P. R., & Wang, M. Q. (1994). Toward interface design for human language technology: Modality and structure as determinants of linguistic complexity. *Speech Communication*, 15, 283-300.
- Oviatt, S. L., DeAngeli, A., & Kuhn, K. (1997). Integration and synchronization of input modes during multimodal human-computer interaction. *Proceedings of Conference on Human Factors in Computing Systems (CHI'97)*, 415-422. New York: ACM Press.
- Oviatt, S. L., & Kuhn, K. (1998). Referential features and linguistic indirection in multimodal language. *Proceedings of the International Conference on Spoken Language Processing*, 6, 2339-2342. Sydney, Australia: ASSTA, Inc.
- Oviatt, S. L., & Olsen, E. (1994). Integration themes in multimodal human-computer interaction. In K. Shirai, S. Furui, & K. Kakehi (Eds.), *Proceedings of the International Conference on Spoken Language Processing* (Vol. 2, pp. 551-554). Yokohama, Japan: Acoustical Society of Japan.
- Oviatt, S. L., & Pothering, J. (1998). Interacting with animated characters: Research infrastructure and next-generation interface design. *Proceedings of the First Workshop on Embodied Conversational Characters*, 159-165, Tahoe City, CA.
- Oviatt, S. L., & vanGent, R. (1996). Error resolution during multimodal human-computer interaction. *Proceedings of the International Conference on Spoken Language Processing*, 2, 204-207. Philadelphia: University of Delaware Press.
- Pankanti, S., Bolle, R. M., & Jain, A. (Eds.). (2000). Biometrics: The future of identification [Special issue]. *Computer*, 33(2).
- Papineni, K. A., Roukos, S., & Ward, R. T. (1997). Feature-based language understanding. *Proceedings of the 5th European Conference On Speech Communication and Technology*, 3, 1435-1438. Rhodes, Greece: European Speech Communication Association.
- Pavlovic, V., Berry, G., & Huang, T. S. (1997). Integration of audio/visual information for use in human-computer intelligent interaction. *Proceedings of IEEE International Conference on Image Processing*, 121-124. Washington, DC: IEEE Press.

- Pavlovic, V., & Huang, T. S. (1998). Multimodal prediction and classification on audio-visual features. *AAAI'98 Workshop on Representations for Multi-Modal Human-Computer Interaction*, 55-59. Menlo Park, CA: AAAI Press.
- Pentland, A. (1996, April). Smart rooms. *Scientific American*, 68-76.
- Perrault, C. R. (1990). An application of default logic to speech act theory. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 161-186). Cambridge, MA: MIT Press.
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *IEEE Proceedings*, 267-296.
- Reithinger, N., & Klesen, M. (1997). Dialogue act classification using language models. *Proceedings of Eurospeech'97, Vol. 4*, 2235-2238. Grenoble: European Speech Communication Association.
- Rhynne, J. R., & Wolf, C. G. (1993). Recognition-based user interfaces. In H. R. Hartson & D. Hix (Eds.), *Advances in human-computer interaction* (Vol. 4, pp. 191-250). Norwood, NJ: Ablex.
- Roe, D. B., & Wilpon, J. G. (Eds.). (1994). *Voice communication between humans and machines*. Washington, DC: National Academy Press.
- Rogozan, A., & Deleglise, P. (1998). Adaptive fusion of acoustic and visual sources for automatic speech recognition. *Speech Communication*, 26(1&2), 149-161.
- Roth, S. F., Mattis, J., & Mesnard, X. (1991). Graphics and natural language as components of automatic explanation. In J. W. Sullivan & S. W. Tyler (Eds.), *Intelligent user interfaces* (pp. 207-239). New York: ACM Press.
- Rubin, P., Vatikiotis-Bateson, E., & Benoit, C. (1998). Special issue on audio-visual speech processing. *Speech Communication*, 26(1&2).
- Rudnick, A., & Hauptmann, A. (1992). Multimodal interactions in speech systems. In M. Blattner & R. Dannenberg (Eds.), *Multimedia interface design* (pp. 147-172). New York: ACM Press.
- Sadek, D. (1991). Dialogue acts are rational plans. *Proceedings of the ESCA/ETRW Workshop on the Structure of Multimodal Dialogue*, 1-29. Maratea, Italy: European Speech Communication Association.
- Sag, I. A., & Wasow, T. (1999). *Syntactic theory: A formal introduction*. Stanford, CA: CSLI Publications.
- Schwartz, D. G. (1993). *Cooperating heterogeneous systems: A blackboard-based meta approach*. Unpublished doctoral thesis, Case Western Reserve University.
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge, England: Cambridge University Press.
- Sejnowski, T., Yuhas, B., Goldstein, M., & Jenkins, R. (1990). Combining visual and acoustic speech signal with a neural network improves intelligibility. In D. Touretzky (Ed.), *Advances in neural information processing systems* (pp. 232-239). Denver.
- Seneff, S., Hurley, E., Lau, R., Pao, C., Schmid, P., & Zue, V. (1998). Galaxy-II: A reference architecture for conversational system development. *Proceedings of the International Conference on Spoken Language Processing*, 1153. Sydney, Australia: ASSTA, Inc.
- Shaikh, A., Juth, S., Medl, A., Marsic, I., Kulikowski, C., & Flanagan, J. (1997). An architecture for multimodal information fusion. *Proceedings of the Workshop on Perceptual User Interfaces (PUI'97)*, 91-93. Banff, Canada.

- Spiegel, M., & Kamm, C. (Eds.). (1997). Interactive voice technology for telecommunications applications (IVTTA'96) [Special issue]. *Speech Communication*, 23(1&2).
- Stork, D. G., & Hennecke, M. E. (Eds.). (1995). *Speechreading by humans and machines*. New York: Springer-Verlag.
- Suhm, B. (1998). *Multimodal interactive error recovery for non-conversational speech user interfaces*. Unpublished doctoral thesis, Fredericiana University, Germany.
- Suhm, B., Myers, B., & Waibel, A. (1996). Interactive recovery from speech recognition errors in speech user interfaces. *Proceedings of the International Conference on Spoken Language Processing*, 2, 861-864. Philadelphia: University of Delaware Press.
- Sun Microsystems. (1998). *Java speech API*. Palo Alto, CA: Author. <http://java.sun.com/products/java-media/speech/>
- Sutton, S., & Cole, R. (1997). The CSLU Toolkit: Rapid prototyping of spoken language systems. *Proceedings of UIST'97: The ACM Symposium on User Interface Software and Technology*, 85-86. New York: ACM Press.
- Turk, M., & Robertson, G. (Eds.). (2000). Perceptual user interfaces [Special issue]. *Communications of the ACM*, 43(3).
- Vergo, J. (1998). A statistical approach to multimodal natural language interaction. *Proceedings of the AAAI'98 Workshop on Representations for Multimodal Human-Computer Interaction*, 81-85. Madison, WI: AAAI Press.
- Vo, M. T. (1998). *A framework and toolkit for the construction of multimodal learning interfaces*. Unpublished doctoral thesis, Carnegie Mellon University, Pittsburgh, PA.
- Vo, M. T., Houghton, R., Yang, J., Bub, U., Meier, U., Waibel, A., & Duchnowski, P. (1995). Multimodal learning interfaces. *Proceedings of the DARPA Spoken Language Technology Workshop*, 233-237. Barton Creek: Kaufmann. Retrieved from the World Wide Web: <http://werner.ira.uka.de/ISL.publications.html#1995>
- Vo, M. T., & Wood, C. (1996). Building an application framework for speech and pen input integration in multimodal learning interfaces. *Proceedings of IEEE International Conference of Acoustic, Speech and Signal Processing*, 6, 3545-3548. Atlanta: IEEE Press.
- Wagner, A. (1990). Prototyping: A day in the life of an interface designer. In B. Laurel (Ed.), *The Art of human-computer interface design* (pp. 79-84). Reading, MA: Addison-Wesley.
- Wahlster, W., Andre, E., Finkler, W., Profitlich, H.-J., & Rist, T. (1993). Plan-based integration of natural language and graphics generation. *Artificial Intelligence*, 63, 387-427.
- Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., & Lang, K. J. (1989). Phenome recognition using time-delay neural networks. *IEEE Transactions on Acoustic, Speech, and Signal Processing*, 37, 328-339.
- Waibel, A., Vo, M. T., Duchnowski, P., & Manke, S. (1995). Multimodal interfaces [Special issue]. *Artificial Intelligence Review*, 10(3&4).
- Wang, J. (1995). Integration of eye-gaze, voice and manual response in multimodal user interfaces. *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, 3938-3942. IEEE Press.

- Wickens, C. D., Sandry, D. L., & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors*, 25, 227-248.
- Wickens, C. D., Vidulich, M., & Sandry-Garza, D. (1984). Principles of S-R-C compatibility with spatial and verbal tasks: The role of display-control interfacing. *Human Factors*, 26, 533-534.
- Wojcik, R., & Holmback, H. (1996). Getting a controlled language off the ground at Boeing. *Proceedings of the First International Workshop on Controlled Language Applications*, 22-31. Seattle, WA.
- Wolf, C. G., & Morrel-Samuels, P. (1987). The use of hand-drawn gestures for text editing. *International Journal of Man-Machine Studies*, 27, 91-102.
- Wu, L., Oviatt, S., & Cohen, P. (1998). *From members to teams to committee: A robust approach to gestural and multimodal recognition*. Manuscript submitted for publication.
- Wu, L., Oviatt, S., & Cohen, P. (1999). Multimodal integration—a statistical view. *IEEE Transactions on Multimedia*, 1, 334-341.
- Zhai, S., Morimoto, C., & Ihde, S. (1999). Manual and gaze input cascaded (MAGIC) pointing. *Proceedings of the Conference on Human Factors in Computing Systems (CHI'99)*, 246-253. New York: ACM Press.

89067668616



B89067668616A

HUMAN-COMPUTER INTERACTION

Contents of Volume 15, Number 4

2000

ARTICLES

- Designing the User Interface for Multimodal Speech
and Pen-Based Gesture Applications: State-of-the-Art
Systems and Future Research Directions** 263
*Sharon Oviatt, Phil Cohen, Lizhong Wu, John Vergo,
Lisbeth Duncan, Bernhard Suhm, Josh Bers, Thomas Holzman,
Terry Winograd, James Landay, Jim Larson, and David Ferro*
- Troubles With the Internet: The Dynamics of Help at Home** 323
*Sara Kiesler, Bozena Zdaniuk, Vicki Lundmark,
and Robert Kraut*
- Prototyping Praxis: Constructing Computer Systems
and Building Belief** 353
Douglas Tudhope, Paul Beynon-Davies, and Hugh Mackay
- Acknowledgment to Reviewers 385
- Author Index for Volume 15, 2000 386
-