Just Do What I Tell You: The Limited Impact of Instructions on Multimodal Integration Patterns

Sharon Oviatt, Rachel Coulston, and Rebecca Lunsford

Center for Human-Computer Communication,
Department of Computer Science & Engineering,
Oregon Health & Science University, 20000 NW Walker Road,
Beaverton, OR 97006, USA +1 503-748-1342
{oviatt, rachel, rebeccal}@cse.ogi.edu

Abstract. Large individual differences have been documented among users in their multimodal integration patterns, which suggest that new user-adaptive approaches to multimodal fusion may be opportune. Before pursuing such an approach, this study explores whether people can be successfully encouraged to switch their multimodal integration pattern to one that is easier to process through the use of explicit instructions. Longitudinal data were collected from young and elderly adults as they used speech and pen input with a simulated map system. Results revealed that only 37% of users switched their integration pattern and maintained it, whereas another 19% never switched their natural pattern and 31% switched but then reverted during a follow-up session. In addition, significant destabilization of elderly users' integration pattern was one "cost" of attempting to instruct a change in pattern. This research underscores the need for user-centered design in future multimodal system development, especially for vulnerable users such as the elderly.

1 Introduction

Research has documented large individual differences among users in multimodal integration patterns, with greater variability among the elderly than younger adults [6,9]. However, state-of-the-art multimodal systems still are based on *fixed temporal thresholds* to determine when modality fusion is "legal" [1]. These temporal thresholds are used to resolve when a person's input is unimodal versus multimodal, and also when sequential signals separated by a lag should be fused into one multimodal interpretation. Since multimodal systems based on fixed temporal thresholds are inaccurate for many users and do not permit tailoring to handle departures from modal patterns, one key direction for future multimodal interfaces is the development of a new class of *adaptive temporal thresholds* that can *detect and adapt to a user's dominant multimodal integration pattern*. Adaptive thresholds are expected to reduce system response delays to approximately 44% of what they currently are for fixed thresholds, and also to significantly improve the synchrony of user-system interaction and overall system reliability.

As a new class of adaptive multimodal interfaces begins to be prototyped, engineers might reasonably ask whether users can't just be trained to deliver their multi-

L. Ardissono, P. Brna, and A. Mitrovic (Eds.): UM 2005, LNAI 3538, pp. 261 – 270, 2005.

modal commands in a simultaneously integrated manner. This could greatly simplify the development of temporal constraints that are needed to build new time-sensitive multimodal architectures. The present research explores this theme of the potential malleability of users' multimodal integration patterns, as well as examining possible differences between younger adults and the elderly.

1.1 Related Research on Multimodal Integration Patterns

Recent research has revealed an unusual bimodal distribution of multimodal integration patterns when users interact with computers. As illustrated in table 1, studies conducted with users across the lifespan have indicated that individual child, adult, and elderly users all adopt either a predominantly *simultaneous* or *sequential* integration pattern during production of speech and pen multimodal constructions [6,8,9]. During a simultaneous integration, speech and pen input is at least partly overlapped in time, whereas during a sequential construction one input mode begins and ends before the second starts. A user's dominant integration pattern can be identified almost immediately, typically on the very first multimodal command, and it remains highly consistent (88-97%) throughout an interactive computer session [6,8,9]. Interestingly, large individual differences and within-subject stability likewise have been documented in the perception of multisensory synchrony [4,7].

Table 1. Percentage of simultaneously-integrated multimodal constructions (SIM) versus sequentially-integrated constructions (SEQ) for children, adults, and seniors

Children			Adults			Seniors		
U	**SIM**	**SEQ**	**User**	**SIM**	**SEQ**	**User**	**SIM**	**SEQ**
SIM integrators:			**SIM integrators:**			**SIM integrators:**		
1	100	0	1	100	0	1	100	0
2	100	0	2	94	6	2	100	0
3	100	0	3	92	8	3	100	0
4	100	0	4	86	14	4	97	3
5	100	0	**SEQ integrators:**			5	96	4
6	100	0	5	31	69	6	95	5
7	98	2	6	25	75	7	95	5
8	96	4	7	17	83	8	92	8
9	82	18	8	11	89	9	91	9
10	65	35	9	0	100	10	90	10
SEQ integrators:			10	0	100	11	89	11
11	15	85	11	0	100	12	73	27
12	9	91				**SEQ integrators:**		
13	2	98				13	1	99
						Non-dominant integrators:		
						14	59	41
						15	48	52
Average Consistency 93.5%			**Average Consistency 90%**			**Average Consistency 88.5%**		

In many respects, these data on individual differences in multimodal integration patterns present an ideal opportunity for adaptive processing, since users are divided into two basic types, with early predictability and high consistency in their integration pattern. Furthermore, recent work has indicated that users' natural dominant integration pattern, whether simultaneous or sequential, spontaneously remains stable over extended time periods [5]. In addition, their dominant integration pattern is resistant to change even when a strong training contingency is delivered [6].

1.2 Goals of This Study

The present research assesses whether an individual's natural dominant multimodal integration pattern, either simultaneous or sequential, can be changed via explicit instructions. It also examines whether this changed pattern then remains stable during a longitudinal follow-up one month later. It was predicted that: (1) most users would switch their integration pattern if explicitly instructed to do so (from simultaneous to sequential, or vice versa), with younger adults more likely than elderly ones to change. One related goal was: (2) to investigate how gradual or abrupt this change in patterns would be as people consolidated their new integration pattern. During the longitudinal follow-up, it was hypothesized that: (3) some users would revert to their natural integration pattern, with elderly users more likely to do so than younger ones due to memory limitations and greater individual differences [2,3]. It also was expected that: (4) elderly users would be more likely to report having forgotten the original instruction than younger ones. Finally, it was hypothesized that: (5) one by-product or "cost" of attempting to change people's natural integration pattern, whether this attempt was successful or not, would be to destabilize or reduce the overall consistency of their pattern.

The long term goal of this research is the development of empirically-based models on users' multimodal integration patterns, which will be needed for deriving optimal temporal thresholds for signal fusion in a new generation of time-sensitive multimodal architectures. One expected outcome of such work is the design of high-performance multimodal systems that are capable of adapting to a full spectrum of diverse users, thereby supporting more tailored and robust multimodal systems.

2 Methods

2.1 Participants, Task and Procedure

There were 16 participants, 6 elderly adults 66-89 years of age, and 10 younger adults 18-61 years of age. Among the elderly 4 were female and 2 male, whereas the young adults included 6 females and 4 males. All participants were native English speakers and paid volunteers and represented varied professional backgrounds. They were healthy and physically active, had no major physical limitations or cognitive impairments, and were not on medications known to influence speech or motor performance.

Participants were instructed to act as volunteers assisting during a flood management exercise. They used a simulated multimodal map system. Instructions from headquarters were displayed as text near the bottom of their screen. The experimenter gave them instructions and practice until they were ready to work. Participants were told that they could use speech and pen input in any way they wished, as long as they used both modalities for each task. The experimenter's instructions initially were unbiased with respect to how users could integrate modalities. Then the experimenter left the room, and the participant completed the first 10 tasks, which constituted an identification band to determine their natural dominant multimodal integration pattern.

Instructional manipulation – After determination of the user's integration pattern, the experimenter then reentered the room to check on the user, and explained that she had forgotten to mention that the system would work best if speech and pen input were integrated together/separately (i.e., whichever was not their natural dominant pattern). This was done without making reference to the user's pattern. If the volunteer had just been identified as a sequential integrator, then he was instructed to provide speech and pen *together, or in an overlapped way*. However, if he was identified as simultaneous, he was instructed to complete one input mode before starting the second so they *would not be presented at the same time*. If the volunteer asked, he was told that it did not matter which mode was used first. To one subject, for example, whose dominant integration pattern was simultaneous, the experimenter said "I forgot to tell you something. I was talking to the programmer, and the system actually works best if you give your spoken and pen input not overlapping; so do one, then do the other. It doesn't matter which one you do first." Following this instructional intervention, the experimenter watched while the participant completed one task correctly as instructed. Then she left the room and the participant completed another 82 tasks while working alone for about one hour.

Longitudinal follow-up – One month after their initial session, each participant returned for a second one. After a brief reorientation and practice using the same system, they completed 83 tasks while working alone for about one hour. On this visit, no instructions were given on how to integrate speech and pen, and no mention was made of the previous instructions they received during their earlier session.

Post-experimental Interview – An oral interview was conducted by the experimenter at the end of the second longitudinal session to determine whether participants were aware of their own integration pattern, and whether they recalled the experimenter's instructions. Participants also were debriefed on the purpose of the study, and it was confirmed that everyone believed they were interacting with a fully functional system.

Simulation Technique – Data collection was accomplished using a dual-wizard high-fidelity semiautomatic simulation technique, as described in previous work [6], with a simulated recognition error rate of 20% throughout each session.

2.2 Research Design, Data Capture and Coding

The research design involved an initial identification band phase, during which a user's dominant multimodal integration pattern was classified based on the first 10 tasks. After this, the user received the main instructional manipulation, which entailed giving different instructions depending on what the user's dominant pattern was, as described above. The main data collection phase followed, with a second longitudinal session following one month later. The experimental design involved two within-subject factors: (1) Longitudinal session (first, second) and (2) Age group (younger, elder). Both sessions and the interviews were videotaped, and multimodal integration patterns, consistency levels, and self-report data then were analyzed as such:

Integration Pattern Classification– All multimodal constructions were classified as either simultaneous or sequential in their temporal integration pattern. A multimodal construction was *simultaneous* if the gesture and speech components were executed with any portion temporally overlapping. A multimodal construction was considered *sequential* if the gesture and speech contained no overlap, and instead a temporal lag was present between the modes. Based on a subject's percentage of simultaneous versus sequential integration patterns for a given session, that person's session also was classified as either a simultaneous (M) or sequential (Q) dominant integration pattern if 60% or more of the constructions for that session represented that pattern, and as non-dominant (ND) if the 60% criterion was not reached for either pattern (i.e., falling between 40-60% consistency range). The dominant pattern identified during a person's baseline identification band (i.e., before instructions) was termed their *natural integration pattern*. If a person responded to instructions by reversing their integration pattern, this was termed the *instructed integration pattern*.

Integration Pattern Change Score– For a given participant, change scores were calculated between the subject's percentage of constructions delivered in their natural integration pattern as defined during the identification band, and the percentage of constructions delivered in that same pattern on a later session following instructions (i.e., either session 1 or 2). This change score was computed by subtracting the percentage on the later session from the original baseline (e.g., %ID - %Session 1).

Integration Pattern Consistency– The percentage of a person's total constructions delivered in their dominant integration pattern was calculated for a given session.

Post-Experimental Interview – Participants' responses to interview questions were summarized as a percentage within each category.

2.3 Reliability

The dual-wizard simulation technique permitted real-time identification and logging of users' integration pattern throughout the session, and previous analyses have revealed it to be 99% accurate when compared with hand codings [9]. Participants' multimodal integration pattern during the baseline identification band was hand verified. The integration pattern for all multimodal constructions throughout the rest of the participants' sessions also was calculated by measuring the start and end of both the speech and pen signal, and programmatically identifying overlap (simultaneous) or lag (sequential) between the two signals. Measurements of start and end of the speech and pen signals were compared between coders for 6% of the data in the first session, and over 80% matched to within 0.1 second.

3 Results

Analyses were based on longitudinal data from approximately 2770 multimodal constructions, including 1740 from younger adults and 1030 from elderly adults.

3.1 Changes in Dominant Integration Pattern

Of the sixteen subjects, all displayed a clear dominant multimodal integration pattern during their identification band - their first ten constructions. Fourteen were identified as simultaneous integrators, and two as sequential. Following the explicit instructional manipulation, twelve subjects or 75% reversed their integration pattern during session 1 as expected. One dropped below the 60% threshold for dominance, but did not actually switch to the reverse pattern completely as instructed. The remaining three subjects never changed their dominant integration pattern at all. These data are summarized in table 2. Of the elder adults, four out of six (67%) switched integration patterns but two did not (33%), whereas eight out of ten younger adults switched (80%) and only two did not (20%). An analysis of the integration pattern change score between the ID band and session 1 for younger versus elder adults revealed that these two groups were not significantly different by Wilcoxon Rank Sum test, $z < 1$, N.S., one-tailed.

Table 2. Dominant multimodal integration pattern and percentage consistency for each elder or younger adult during their ID band, first session, and second follow-up session, with consistency levels broken down by first versus second half for the main sessions [M-Simultaneous; Q-Sequential; ND-No dominant pattern]

	ID Band	**Session 1**			**Session 2**		
Subject Age	**Natural Dominant Pattern**	**Dominant Pattern (Overall)**	**1st half**	**2nd half**	**Dominant Pattern (Overall)**	**1st half**	**2nd half**
79	M (100.0%)	Q (81.7%)	82.9%	80.5%	M (80.3%)	85.7%	75.6%
71	M (88.9%)	Q (95.1%)	97.6%	92.7%	M (61.4%)	81.0%	41.5%
78	M (100.0%)	Q (66.3%)	*57.1%*	75.6%	M (91.6%)	85.7%	97.6%
66	M (80.0%)	Q (81.5%)	78.6%	84.6%	Q (89.2%)	83.3%	95.1%
69	M (90.0%)	M (83.8%)	81.1%	86.5%	M (96.2%)	97.4%	94.9%
89	M (100.0%)	M (67.5%)	73.8%	61.0%	M (85.2%)	85.0%	85.4%
Elder Avg	**93.2%**	**79.3%**	**78.5%**	**80.1%**	**84.0%**	**86.4%**	**81.7%**
26	M (100%)	Q (86.7%)	88.1%	85.4%	*ND (56.6%)M*	42.9%	70.7%
61	M (70%)	Q (79.5%)	83.3%	75.6%	M (60.0%)	42.5%	77.5%
45	M (100%)	Q (82.9%)	78.0%	87.8%	M (97.6%)	95.2%	100.0%
27	M (100%)	Q (85.0%)	80.5%	89.7%	Q (97.3%)	100.0%	95.1%
19	M (80%)	Q (98.8%)	97.6%	100.0%	Q (98.8%)	100.0%	97.6%
23	M (100%)	Q (100.0%)	100.0%	100.0%	Q (100.0%)	100.0%	100.0%
18	M (100%)	Q (94.0%)	88.1%	100.0%	Q (100.0%)	100.0%	100.0%
53	Q (70%)	M (91.6%)	85.7%	97.6%	M (91.5%)	87.8%	95.1%
33	M (100%)	M (96.4%)	97.6%	95.1%	M (98.8%)	97.6%	100.0%
19	Q (70%)	*ND (51.25%)Q*	*55.0%*	*47.5%*	M (100.0%)	100.0%	100.0%
Younger Avg	**89.0%**	**86.6%**	**85.4%**	**87.9%**	**90.1%**	**86.6%**	**93.6%**
Overall Avg	**90.6%**	**83.9%**	**82.8%**	**85.0%**	**87.8%**	**86.5%**	**89.1%**

By the second longitudinal session one month later, figure 1 illustrates that only six of the twelve participants who had initially switched their dominant integration pattern (i.e., 6 of original 16) continued with this instructed pattern, or just 37%. Five others reverted back to their natural integration pattern (i.e., 5 of 16), or 31%, and the remaining person reverted back partially but fell within the non-dominant zone. Of the 3 participants who had maintained their natural integration pattern during the first session (i.e., 3 of 16), all or 19% continued to maintain this pattern during the longitudinal follow-up. The last person, who had partially switched patterns during the first session but remained within the non-dominant zone, eventually did switch to the instructed pattern in a delayed manner during the second session.

With respect to differences between elder and younger adults on the longitudinal follow-up, three of the four elders, or 75% of those who had switched to the instructed pattern during the first session reverted back to their natural pattern. In contrast, only two of the eight younger adults, or 25% who had switched, reverted back. As shown in figure 2, by the second session, six out of ten younger adults were displaying the instructed integration pattern as their dominant one (60%, as compared with 80% on session 1), whereas only one of six elderly adults did so (16.5%, as compared with 67% on session 1). An analysis of the integration pattern change score between the ID band and session 2 for younger versus elder adults revealed a significant difference between the two groups by Wilcoxon Rank Sum test, z = 1.37, $p < 0.05$, one-tailed.

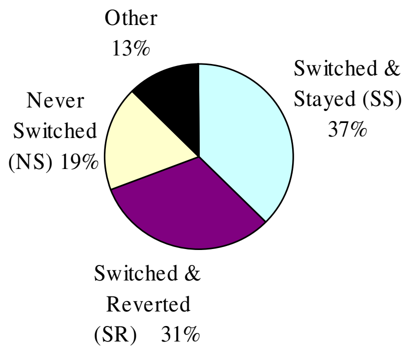

Fig. 1. Percentage of all participants who switched to the instructed integration pattern and maintained it through the longitudinal follow-up (SS), switched but reverted back to their natural integration pattern (SR), or never switched from their natural integration pattern (NS)

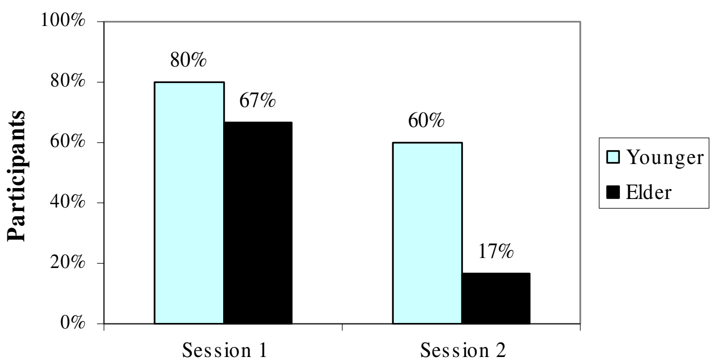

Fig. 2. Percentage of younger versus older adults who displayed instructed integration pattern on session 1 and session 2

3.2 Changes in Consistency of Integration Pattern

Table 2 (see preceding page) summarizes participants' dominant integration pattern (M, Q, ND) and also their average consistency level during the identification band, session 1, and session 2. Since one hypothesis was that the instructional intervention itself might destabilize people's integration pattern and result in reduced consistency of this pattern, this was evaluated in the results. As shown in table 2, participants' average consistency level during the identification period was 90.6%, which decreased to 82.8% *immediately following instructions* in the first half of session 1, a significant drop by Wilcoxon Signed Rank test, T+ = 90, N = 15, $p < 0.05$, one-tailed. A follow-up within-subject comparison by age group revealed that elderly adults' average integration pattern consistency decreased from 93.2% to 78.5% between the identification band and first half of session 1, which was a significant drop by Wilcoxon Signed Rank test, T+ = 19, N = 6, $p < 0.05$, one-tailed. However, younger adults' decrease in consistency from 89.0% to 85.4% for this same interval was not a statistically significant one, Wilcoxon Signed Rank test, T+ = 28, N = 9, NS, one-tailed. Figure 3 illustrates this difference between elder and younger adults in the destabilization of their integration pattern immediately following instructions.

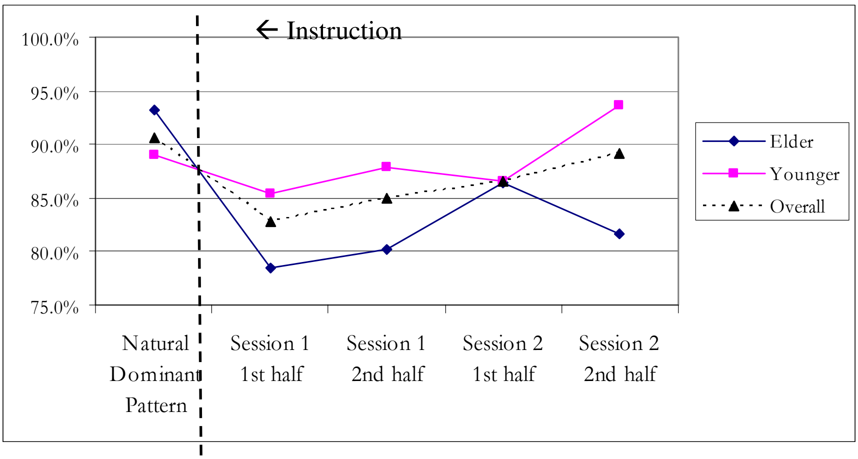

Fig. 3. Progressive changes over time in younger versus elder adults' average integration pattern consistency level from their initial baseline period through the end of session 2

To assess whether this drop in consistency had returned to the baseline level by the end of session 2, or whether participants' integration pattern consistency had recovered by this time, further comparisons were conducted on their average consistency level during the identification period (90.6%) versus the *second half of session 2* (89.1%), which no longer represented a significant decrease by Wilcoxon Signed Rank test, T+ = 39, N = 12, N.S., one-tailed. A follow-up comparison of elderly adults' average integration pattern consistency during the identification period (93.2%) versus the *second half of session 2* (81.7%), revealed that the gap they originally displayed did not continue to reflect a significant decrease one month later, Wilcoxon Signed Rank test, T+ = 15, N = 6, N.S., one-tailed. A similar comparison on younger adults for this same time interval revealed that their average consistency had returned from 89.0% back up to 93.6%, which actually exceeded their original level. Figure 3 also illustrates this difference between elder and younger adults in the return of stability in their integration pattern by the end of session 2.

3.3 Self-report on Integration Patterns

All subjects reported being aware of their integration pattern, although only 12 of the 16 (75%) were correct in reporting whether it was mainly simultaneous or sequential.

The other four (25%) self-reported the wrong integration pattern. When asked whether their integration pattern had changed between the first and second sessions, six people (38%) reported correctly but another nine (56%) were wrong about whether their integration pattern changed and one (6%) did not remember.

When asked whether they remembered the instruction in session 1 about how to integrate the two input modes, 9 of 16 subjects (56%) correctly recalled that the experimenter had asked them to change their integration pattern and what the instruction had been. The other 7 (44%) either did not remember receiving an instruction at all or failed to remember what it was. In addition, for 63% of participants the ability to recall instructions correctly corresponded with having switched and/or maintained their integration pattern accordingly by session 2. Among younger adults, 50% correctly recalled the original instruction, although 67% of the elderly actually remembered it, which clearly did not support the interpretation of greater selective forgetting among the elderly group.

4 Discussion

This research underscores that future multimodal systems need to accurately model users' existing natural integration patterns, rather than naively assuming that instructions can prompt users to adopt a particular style that may be easier for the system to process. Only 37% of users in this study switched their natural integration pattern to the instructed one and maintained it one month later, whereas 19% never switched their natural pattern at all, and another 31% switched but then reverted back to their natural pattern during the follow-up session. That is, permanent switching was uncommon in spite of the fact that all users had demonstrated their understanding of the experimenter's original instruction by composing an appropriately integrated construction. However, the self-report data indicated that 25% of users were not aware of what the temporal organization of their multimodal communication had been over the last hour, and 62% of users either failed to recall or incorrectly remembered whether their integration pattern had changed from session 1 to 2. These data on users' limited awareness of their multimodal integration patterns at least partially explain why explicit instructions were so ineffective in prompting the desired change.

As illustrated in figure 2 and confirmed by the analysis of younger and elderly adults' change scores between their baseline and session 2, elderly users also were significantly less likely than younger ones to switch and maintain the instructed integration pattern. By the longitudinal follow-up, only 16.5% of the elderly still maintained the instructed pattern, whereas 60% of younger adults did so. Perhaps surprisingly, self-report data did not support the interpretation that the elderly forgot the instructions more frequently than younger adults. They simply were more likely to either persist or to revert back to their natural dominant pattern. In addition, a significant temporary destabilization of participants' integration pattern was a *cost* of attempting to instruct them to change their natural pattern, as shown in figure 3. There was a significant drop in elderly adults' average consistency level from 93.2% during their baseline period to 78.5% immediately following instructions on the first half of session 1. By the second half of the longitudinal follow-up, their consistency level had climbed back up to 81.7%, which no longer represented a significant departure from baseline. Although younger adults were not completely immune from the destabilizing impact of instructions, they nonetheless were not significantly disrupted by them to the same extent as the elderly.

Taken together, these results underscore that a user-centered design perspective is needed to guide successful multimodal system development, since (1) the majority of users cannot be expected to change their natural multimodal integration pattern to suit system processing capabilities, and (2) attempts to change their pattern incurs a cost by destabilizing user-system interaction. The present data also clarify that user-centered design is more critical for elderly users, since they are less likely to adapt to the system, and also are more adversely affected by attempts to instigate change in their interaction patterns. These results have implications for the effective design of "aging-in-place" and other emerging elder interfaces.

The long-term goal of this research is the development of empirically-based models on users' multimodal integration patterns. Such models will be needed for deriving optimal temporal thresholds for signal fusion for a new generation of time-sensitive multimodal architectures. One outcome of such work will be the design of high-performance multimodal systems that are capable of adapting to a full spectrum of diverse users, thereby supporting more tailored and robust multimodal systems.

Acknowledgements

Thanks to B. Xiao and K. Hollingshead for data collection, scoring and programming, to M. Wesson for acting as wizard, and to members of CHCC for insightful discussions. This research was supported by DARPA Contract No. NBCHD030010 and NSF Grant No. IIS-0117868.

References

1. Cohen, P.R., Johnston, M., McGee, D.R., Oviatt, S., Pittman, J., Smith, I., Chen, L., Clow, J.: QuickSet: Multimodal interaction for distributed applications. Proc. Multimedia'97 (1997) 31-40
2. Craik, F.I.M., Salthouse, T.A., (eds.): Handbook of Aging and Cognition. 2 ed. LEA, Mahwah, NJ (2000)
3. Czaja, S.J., Lee., C.C.: Designing computer systems for older adults. In: J. Jacko, A. Sears, (eds.): Handbook of Human-Computer Interaction. LEA, New York (2003) 413-427
4. Mollon, J.D., Perkins, A.J.: Errors of judgement at Greenwich in 1796. Nature, Vol. 380. (1996) 101-102
5. Oviatt, S., Lunsford, R., Coulston, R.: Individual differences in multimodal integration patterns: What are they and why do they exist? Proc. Human-Factors in Computing Systems (CHI) (2005) in press
6. Oviatt, S.L., Coulston, R., Tomko, S., Xiao, B., Lunsford, R., Wesson, M., Carmichael, L.: Toward a theory of organized multimodal integration patterns during human-computer interaction. Proc. ICMI (2003) 44-51
7. Stone, J.V., Hunkin, N.M., Porrill, J., Wood, R., Keeler, V., Beanland, M., Port, M., Porter, N.R.: When is now? Perception of simultaneity. Proc. Royal Society: Biological Sciences, Vol. 268. (2001) 31-38
8. Xiao, B., Girand, C., Oviatt, S.L.: Multimodal integration patterns in children. Proc. ICSLP (2002) 629-632
9. Xiao, B., Lunsford, R., Coulston, R., Wesson, M., Oviatt, S.L.: Modeling multimodal integration patterns and performance in seniors: Toward adaptive processing of individual differences. Proc. ICMI (2003) 265-272