A Simple Index for Multimodal Flexibility

Antti Oulasvirta, Joanna Bergstrom-Lehtovirta
Helsinki Institute for Information Technology HIIT
Aalto University and University of Helsinki
firstname.lastname@hiit.fi

ABSTRACT

Most interactive tasks engage more than one of the user's exteroceptive senses and are therefore multimodal. In real-world situations with multitasking and distractions, the key aspect of multimodality is not which modalities can be allocated to the interactive task but which are free to be allocated to something else. We present the *multimodal flexibility index* (MFI), calculated from changes in users' performance induced by blocking of sensory modalities. A high score indicates that the highest level of performance is achievable regardless of the modalities available and, conversely, a low score that performance will be severely hampered unless all modalities are allocated to the task. Various derivatives describe unimodal and bimodal effects. Results from a case study (mobile text entry) illustrate how an interface that is superior to others in absolute terms is the worst from the multimodal flexibility perspective. We discuss the suitability of MFI for evaluation of interactive prototypes.

Author Keywords

Multimodal interaction, modality allocation, mobile human-computer interaction, attention, multitasking.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Design, Human factors.

INTRODUCTION

Almost all human–computer interaction (HCI) situations engage more than one human sense and can therefore be considered *multimodal*. In any given situation, some modalities are not engaged at all, some are but only more passively, and others must be actively deployed if the sensorimotor capacities are to be positioned optimally for action and feedback with the interface. Even the seemingly simple act of pressing a button on a mobile device in fact involves coordination of multiple modalities (see e.g., [23]): simultaneous proprioceptive and visual feedback for coordinating the finger's movement as it closes on the button, mechanoreceptors firing as the button is pressed, visual registration of a character appearing on the display, and hearing of a beep as the system's feedback.

In the development of "multimodal interfaces," researchers have traditionally devoted effort to the question of how to orchestrate sensorimotor capacities optimally for interaction with an interface. Multimodality as viewed in this context could be termed "intra-interface multimodality." In this paper, we turn the question upside down: *what modalities are available to be allocated to tasks other than the current one the user is engaged in?* This question, of "extra-interface multimodality," is a timely one, particularly in the area of mobile HCI [3, 12, 17, 22, 27]. For example, if, during writing of a text message, something happens that causes distraction or reallocation of a sensory modality—e.g., someone asks for directions, a cyclist suddenly approaches, or it is so cold that the fingers start freezing—will you still be able to finish the message without significant costs to performance? We believe that the flexibility of allocation is important whenever there are 1) secondary tasks, distractions, or changes in multitasking strategy; 2) environmental factors such as noise, light, smell, or vibration; or 3) physiological changes that lower transduction capacity (for example, due to brightness or low temperature).

Our initial motivation to study this issue came from the casual observation that two interfaces that *nominally* involve the same sensory modalities may be very different in how well they allow modalities to be employed simultaneously for something else.

To address this, we operationalize the *multimodal flexibility index* (MFI) of a task as the average of performance changes measured over conditions in which the sensory modalities to be studied are blocked. The magnitude of change in user performance caused by blocking is a quantitative indicator of a task's "dependency" on the blocked modality. Intuitively, MFI denotes user ability to reach high performance despite modality withdrawals. The index will be 1 when the highest level of performance is reached in all blocking conditions. If the blocking of one modality decreases performance, the index also decreases. Its value will be 0 if performance in all blocking conditions is at floor level. This corresponds to the situation wherein the user must stop everything else and allocate modalities to the task, or cannot operate the system because a modality is not available. Various derivatives are

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.

provided to detail the dependence of performance on particular modalities. The method can be applied with two to an arbitrarily large number of modalities, though four is the practical maximum.

To study the method's usefulness and practical implementation, we carried out an experiment comparing three input interfaces for mobile devices (using a 12-key keypad and a full QWERTY keyboard with physical or touchpad keys) that nominally engage the same set of senses. The results confirmed our initial observation and indicated that the interface that was the best in *absolute* performance was the worst from the multimodal flexibility viewpoint. Derivative indices allowed us to analyze the situation in more detail, considering possible causes for the differences observed.

We conclude the paper by discussing the method's limitations and potential. On the positive side, the method

1. captures a wide range of outcomes in a single study
2. offers a precise meaning for multimodal flexibility
3. is reasonably cost-efficient—in our study, running one subject took about an hour and the blockings could be administered with inexpensive materials
4. enables practitioners, who may not have time to work with theories to predict the outcomes of complex situations (we could not predict the outcomes of our study), to get quick feedback.

We discuss opportunities for future work and good practices that can mitigate the most probable problems in applying the method. Our current recommendation is to utilize the method in a comparative setting before committing to a particular design solution, and before one conducts more expensive field studies or embarks on cognitive modeling. This method therefore complements the others available by offering developers a way to get quick feedback in early stages of design.

BACKGROUND

In HCI, the term "multimodality" often refers to novel input techniques or communication channels (see [16]). In this paper, *modality* refers to human exteroceptive senses. Because the method assumes no theory of what those modalities are or how they operate, we seek no further definition either—although we do take a stance in the study. Within this scope, we are interested in how freely the user can allocate a modality or its capacities away from the interactive task.

The Idea of Blocking

Technically, MFI is *a method of residues*: "Subduct from any phenomenon such part as is known by previous inductions to be the effect of certain antecedents, and the residue of the phenomenon is the effect of the remaining antecedents" [8: p. 264]. The index is calculated on the basis of data collected from blockings of *all* combinations of the modalities under scrutiny. *Blocking* a modality here means that the experimenter administers something to distract or prevent a modality from being fully deployed in the task. The logic of blocking is to emulate the situation wherein a resource is fully or partially allocated away or inoperable for some other reason, such as physiological changes.

This idea is rooted in cognitive psychological studies of cooperation and interference among cognitive faculties. While the Gestalt psychologists had already employed the methods of subtraction and residues, an important systemization of the techniques was given in Posner's [31] book *Chronometric Explorations of Mind*, where he analyzed the time course of human information processing pathways. Numerous papers have been published that utilize blocking to study bimodal performance. Illustrating a typical blocking study, to examine the cooperation of modalities in object recognition [7], subjects were given a wooden object in one hand and asked to state whether its shape is the same as that of a slot that was touched with the other hand. This was done in three conditions—visually only, haptically only, and bimodally with both—and it was found that bimodally performance was better than unimodally.

MFI, too, is based on observations of changes that are compared to non-manipulated ("none-blocked") baselines. The defining difference from psychology is *why* this is done: we are interested not in understanding the modalities themselves but in emulating the situation seen when real-world demands render modalities unable to be allocated fully.

Alternative Techniques

Similar ideas have been explored in many domains of HCI. Six main categories of methods can be recognized. The first two (1–2) are empirical and follow a logic similar to MFI's, the next two (3–4) aim at direct simulation of real-world situations, and the final ones (5–6) are analytical.

1. Visual occlusion [34]**:** Visual occlusion involves temporary occlusion of the visual field—e.g., a driver must press a button to get a brief (say, 1.5 s) glance at the road; the situations in which this is done reflect vision-critical moments of driving. The method partly shares with MFI the idea of blocking human senses. The differences are that 1) the measurements reflect strategic control of modality allocation; 2) the data can better answer the "when" question than the "how much" one; and, 3) moreover, we are not aware of attempts to generalize the method to two or more modalities.

2. Secondary tasks that overload specific cognitive faculties: In this paradigm, an individual performs two tasks simultaneously, one of which is the main task and the other a task designed to selectively knock out or distract a particular cognitive subsystem. Performance is compared to single-tasking baselines. Typical secondary tasks include random number generation to overload the central executive, repetition of meaningless syllables to distract the auditory loop, and imagery tasks to distract the visual capacities of cognition. The main task is argued to involve a particular cognitive subsystem if the corresponding overloading task significantly degrades performance. The main difference is that this method addresses a different source of performance degradation than does MFI: central interference within and between cognitive faculties (see [29, 36]).

3. Dual-tasking with natural secondary tasks, or laboratory analogues of them**:** This allows directly testing the consequences of a particular secondary task for target performance (e.g., [4]). The difference from methods 1 and 2 is that they look at modality involvement *irrespective* of what the particular secondary task is; their results are supposed to generalize to *any* secondary task that recruits a particular resource.

4. Changing the modality of information display/access: In this extensively utilized paradigm, the human exteroceptive sense as such is not affected, only the output or input signal of the user interface (UI) (e.g., [9]). For example, Liu [21] compared visual-only, aural-only, and multimodal display of information during driving, and Burke et al. [6] present a meta-analysis of visual-auditory and visual–tactile feedback, comparing them to visual-only. In a variant of this method, multimodal support is removed entirely or changed adaptively [17]. In this paradigm, what is varied is the modalities that are simultaneously stimulated by the UI, while allocation of sensory modalities is not affected. When the UI is the only source of stimulation of a particular sense, the manipulation is effectively the same as in the MFI method. Jacko et al. [15], for example, studied three types of multimodal feedback for users with macular degeneration, utilizing a combinatory logic that is analogous to our measure of modality-dependency (or "D-value").

5. Task analysis: One may not need experiments to understand the roles that senses play in interactive tasks. Oviatt [28] proposed using task analysis to reveal points where users are more likely to interact multimodally, and Baber and Mellor [1] proposed using critical path analysis for identifying constraints to cooperation of senses in a task. Simple analysis can also be done with the Multiple Resource Theory of Wickens [36], which allows prediction of dual-task interference on the basis of four variables: stages (cognitive vs. response), sensory modalities (auditory vs. visual), codes (visual vs. spatial), and channels (focal vs. ambient). If these methods are applicable in a given case and yield valid predictions, they may obviate the need to carry out experiments. However, while we cannot claim that these methods could not have predicted the results of our study, we suspect that a difficulty would arise in the fact that the three interfaces we tested in our study are very similar from pre-empirical perspectives.

6. Cognitive modeling: Cognitive models such as ACT-R [33] and EPIC [18] have provided the richest description of multimodal flexibility, in the sense that they make empirically testable predictions of performance variables *and* shed light on the underlying processes. Issues relevant to modality allocation, such as task-switching costs, time-sharing strategies, central and peripheral interference, memory retrieval, and control of action, can be analyzed with these models. Cognitive models can also be used for exploratory purposes, through charting of the space of possible interaction strategies for an interface with the related tradeoffs [5].

In view of the sophistication of cognitive models, why would one choose any other method? The reasons are pragmatic. First, cognitive models are criticized for having steep learning curves—they require expertise in model architectures, human cognition, and programming. Second, present-day models may be insufficient for novel HCI situations. The off-the-shelf models do not cover all aspects of modalities, for example, and the number of models of task domains available is limited. Third, a practitioner applying these models may unwillingly have to take stances to controversial theoretical debates—such as whether interference arises from central bottleneck limitations or from graded capacity-sharing [29]. For practitioners, empirical methods have a role in providing initial understanding of a novel situation before one embarks on theory construction and modeling.

OPERATIONALIZATION

The calculation of MFI is based on performance scores obtained over all combinations of blockings.

Let us indicate with the numbers 1, 2, …, n each modality under scrutiny and with M the set of modalities we are interested in; for example, when n=3, M={1,2,3}. Now, the set of blocking conditions B we need for an experiment is $B = \mathcal{P}(M) \setminus \{M\}$; in other words, B is the power set of M, $\mathcal{P}(M)$, from which the none-blocked condition is removed. With the none-blocked baseline, the total number of conditions needed for an experiment is 2^n. For example, were we interested in audition (a) and vision (v), we would need to run an experiment with four conditions: $\emptyset$, a, v, av ($\emptyset$ denotes the "empty" or "all-blocked" condition). Both none-blocked and all-blocked conditions are necessary for MFI. (For example, Table 2 shows the combinations for the three modalities examined in our study.)

For calculation of the index, the performance scores obtained are first normalized per interface for one subject by dividing every score of that user in a blocking condition by the none-blocked score (the baseline). By implication, the means in the none-blocked conditions are always 1 and the other normalized scores range from 0 to 1, where 0 is the floor level that indicates zero performance or failure in the task.

We use the notation s_b to denote the performance score observed in blocking condition b. For example, s_{av} indicates a score recorded when modalities a and v were both available (not blocked) while others (if any) were blocked. Now, the MFI is the mean of the scores in B, as follows.

$$MFI = \frac{\sum_{b \subset B} s_b}{2^n - 1} \qquad (1)$$

For example, if $M=\{a,v\}$ and $s_{\emptyset}=0$, $s_a=2.5$, $s_v=1.0$, and $s_{av}=5.0$, the scaled values would be $s_{\emptyset}=0$, $s_a=0.5$, $s_v=0.2$, and $s_{av}=1.0$, and MFI=$(0+0.5+0.2)/(2^2-1)=0.23$. The interpretation of this result is that, on average, reallocating a modality degraded performance to 23% of the maximum.

This formulation has intuitively appealing properties. First, the index ranges from 0 to 1. Second, the index is not determined by *absolute* performance (although, as we will discuss, it can surreptitiously affect it). Third, statistical testing can be performed on MFI. Fourth, MFI generalizes to any *n*. Since five modalities already yields $2^5=32$ conditions, any more than four (with its 16 conditions) is impractical.

MFI delivers only a single number to describe a complex pattern, and one is likely to need further indices to detail the situation. Below, we define a few indices that tie in with existing work in the field of multimodal performance.

Indices for a Single Modality

From the data, we can calculate "the dependence of performance on modality *m*" as the average of performance changes when *m* is *added* to corresponding conditions wherein *m* was not present:

$$D_m = \frac{\sum_{b\subset B}(s_{b\cup\{m\}} - s_b)}{2^{n-1}} \qquad (2)$$

While the formula looks complex, the idea is simple. In our example, we calculate D_a as the average of changes when audition is added to the condition where it was not present: $D_a=((s_a-s_{\emptyset})+(s_{av}-s_v))/2=(0.5+0.8)/2=0.65$. D_m is interpreted as the average decrease in performance caused by the blocking of a modality. One could calculate an index of *in*dependence by subtracting D_m from 1.

We can make one further demarcation based on D_m: We call an interactive task *m-dependent,* if $D_m > 0.5$. In other words, the removal of *m* yields a drop of 50% in performance (the threshold should, of course, be set on the basis of examination of a particular performance variable).

Using the subtraction method to estimate the importance of a modality is not a new idea in HCI (e.g., [15, 19]). Our contribution is to provide a general formulation for dependency and place it in the context of multimodal flexibility. Our formulation requires the experiment to include the all-blocked condition, which, for example, Jacko et al. [15] did not include. Moreover, as was discussed in the Background section, our aim is not to estimate optimal modalities for feedback but to assess the robustness of performance under conditions where some modalities are not (fully) available.

Indices for Bimodality

Let us now consider the case of two modalities *a* and *b*. A few definitions (see also [30]) can be provided:

- *a* is *complementary* to *b* if $s_{ab} > \max(s_a, s_b)$.
- *a* and b are *synergistic* if $s_{ab} > s_a + s_b$.
- *a* and *b* are *additive* if $s_{ab} = s_a + s_b$.
- *a* and *b* are *interchangeable* if $s_a = s_b = s_{ab}$.
- *a* is *dominant* over *b* if $s_{ab} = s_a > s_b$.
- *a* and *b* are mutually *distractive* if $s_{ab} < \max(s_a, s_b)$.

There are examples of these six outcomes in the literature.

For example, there are numerous examples of intersensory facilitation (e.g., [2, 31]). In the above definitions, synergistic and additive effects can be considered special cases of the complementary effect. The additive effect may be seen in cases where the user can fully allocate a *supportive* function to one modality but nothing else without simultaneously hampering the performance of the other modality.

Interchangeability may occur, when a task can be performed with either of two modalities but simultaneous attention to both modalities is not possible or does not improve performance. Hoggan et al. [13] explored interfaces that can provide tactile and auditory icons with the same information content for a mobile user. If the interface worked perfectly, interchangeability could be observed.

The dominance effect has been suggested to apply for many tasks—for example, dominance of vision in driving [32]. Mutual distraction (or conflict) occurs when the addition of a modality hampers performance. One can, for example, consider the situation wherein poorly designed spatial sounds distract use of vision to the extent that blocking of audition improves performance. This could also result in cases in which an attention shift in one modality causes a shift in another modality (see [2, 10]).

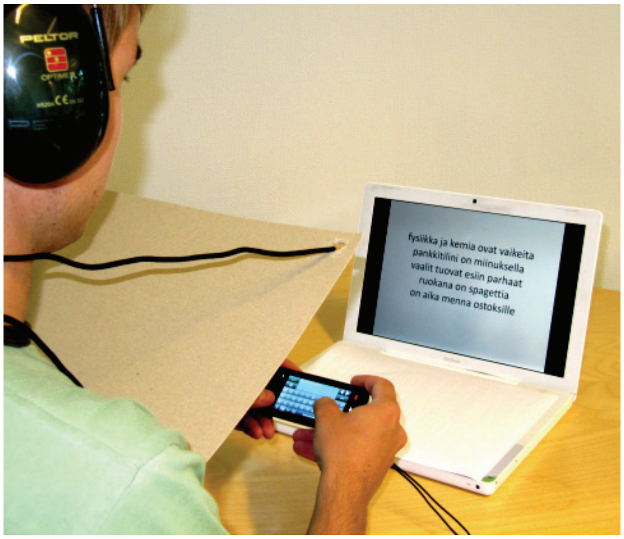
PELTOR
fysiikka ja kemia ovat vaikeita
pankkitilini on miinuksella
vaalit tuovat esiin parhaat
ruokana on spagettia
on aika menna ostoksille

Figure 1. A user in the *t* condition—tactile feedback *un*blocked and vision (cardboard) and audition (ear protection) blocked.

Table 1. Ideas for modality blockings.

Vision (sight)
Turning off the display
Blindfolds or closed eyelids
Pinholes to restrict field of vision
Distorting prisms
Occlusion of the focal area of vision, forcing use of peripheral vision
Color-distorting glasses
Forcing fixation to a fixed location, to distract tracking
Overloading distracters to the device or the environment
Causing abrupt distracting events, to draw attention
Cardboard to occlude the device
User-controllable glasses for visual occlusion
Turning or tilting the user
Audition (hearing)
Turning off the sounds
Ear protection or noise cancellation earphones
Distortive band filters on earphones or in the sound-emitting unit
Imposing temporal delay on sound
Covering one ear to distort binaural spatialization
Causing noise or distracting environmental sounds
Turning or tilting the user
Tactition (touch)
Turning the tactile feedback off
Adhesive surfaces on the device to cover physical boundaries
Distortive (bumpy) surfaces to cover physical boundaries
Surgical gloves or silicone covering of the fingertips
Allowing only one hand to be used
Rotation/turning of objects / the subject to unfamiliar positions
Local anesthesia (dangerous)
Equilibrioception (balance)
Rotating the user rapidly many times
Irritating the outer ear with cool water (dangerous)
Computerized dynamic posturography (CDP)
Proprioception (body position and movement)
Local anesthesia (dangerous)
Forcing a fixed posture
Use of the non-dominant hand
Restricting hand movement to prevent exploration of shapes
Restricting use of the fingers, to force use of only the palm
Adding a secondary motor task
Olfaction (smell)
Covering/closing the nostrils
Imposing distracting odors
Covering the source of odors
Nasal ranger with distortions
Gustation (taste)
(Same as for olfaction)
Substance to cause distortion for chemoreceptors
Alteration of the edible substance

Data Collection Protocol

The collection of data for MFI takes place in an "analogy experiment": the task is carried out in as natural conditions as possible but with no external distractions. Because comparison across tasks and interfaces is problematic (see Discussion), we recommend designing the study as a within-subjects comparative experiment. Now, given a) two or more interface solutions one wants to compare and b) a task, the outline for data collection proceeds as follows:

1. Decide on the modalities that will be blocked. Identification of candidates could be based on user observations or analytical work.
2. Implement blockings. In the study we report on, we sought to employ inexpensive means for blockings, but there are more options suggested by related literature. Our preliminary ideas are listed in Table 1.
3. Develop a dependent variable for performance of the main task that is reliable and sensitive.
4. Ensure comparable conditions, particularly the modalities and interface solutions in different blocking conditions.

The rest of the steps follow standard experimental procedures, with the following precautions:

5. Employ a within-subjects experiment design, counterbalancing the order in which 1) the interfaces and 2) blocking conditions and 3) the tasks (if more than one) appear across subjects.
6. Decide on the level of statistical power desired and calculate the required sample size.
7. Design pre-trial instructions and practice so as to ensure that performance under blocking conditions does not overly reflect the novelty of the situation.
8. After running a pilot, execute the experiment.
9. After preprocessing the data to address outliers and missing data, normalize the scores and calculate the MFI and derivatives. (We provide an Excel sheet for these calculations. Please see Acknowledgements.)

STUDY: MOBILE TEXT INPUT

We chose mobile text input as the study domain, because mobile interaction is known to involve much multitasking [27] and there are efforts to develop interfaces that allow the user to better allocate modalities (e.g., "eyes-free interaction" [3, 22]). We decided to compare three input interfaces that nominally engage the same sensory modalities (see Figure 2):

1. **Touchpad–QWERTY**: The full-QWERTY touchpad of the Nokia XpressMusic 5800
2. **Physical–ITU12**: The ITU E.161 12-key telephone keypad of the Nokia E75
3. **Physical–QWERTY**: The full-QWERTY keyboard of the Nokia E75.

In choosing three modalities for one study instead of two, we hoped to see whether the method scales up from the typical two modalities studied in prior work. The three modalities were chosen in view of what is already known from the literature.

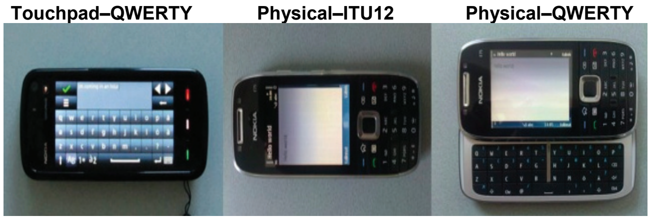

Figure 2. The interfaces compared in the study.

1. *Tactile feedback* to the fingertips was blocked on a) the edges of the buttons and b) button releases, which have been shown to be important for performance in text input and challenged by environmental vibrations (e.g., in a subway environment) and in walking [12, 25]. For implementing this blocking, we explored alternatives from local anesthesia to surgical gloves and silicone-covered fingertips used by clockmakers. However, we ended developing a thin plastic layer attached to the keypads that prevents the user from feeling button releases and the edges of buttons (see Figure 3). While this blocking is imperfect, we were interested not in the absolute performance degradation it caused but in comparing the three interfaces.

2. *Vision* for locating the buttons and coordinating hand movements and for feedback on key presses from the display was blocked with cardboard that occluded the mobile device but did not occlude the line of sight to the task stimuli on a laptop screen (see Figure 1).

3. *Auditory feedback* for button releases and the phone's key-press sounds was blocked by ear protection and by turning off auditory feedback from the device.

Method

Subjects

Twelve students were recruited for the study from a local technical university. Their mean age was 22.8, with an age range of 21 to 26 years (SD=1.6), and the sample was roughly even in gender terms (seven of the subjects were male). As for usage experience, 11 were currently using an ITU keypad, seven with predictive text entry and four without. One subject was using a physical QWERTY keyboard but was also experienced in using an ITU keypad. Two subjects reported that they send fewer than 10 text messages per month, five reported 10–50, four between 50 and 100, and one over 100.

Task and materials

The task was made as similar as possible to that of writing a text message; real words and sentences were used. The task was to type words as correctly as possible for 30 seconds. For every task, five sentences were presented on the computer screen at the same time. The phrases used were from a set of 500 sentences translated into Finnish [14], the subjects' native language, from the original set by MacKenzie and Soukoreff [24]. No special characters, punctuation marks, uppercase letters, or umlauts were used.

Because 30 seconds is too long for memory-based transcription, the sentences remained visible for the duration of the task. Therefore, the task can be considered to involve copying rather than text generation. The copying involved in the task potentially presents a form of multimodal task that differs from text generation: The user has to read the separate computer screen as well as attend to the mobile device.

Apparatus

With the XpressMusic 5800, we used the touchpad–QWERTY interface with horizontal layout. Levels of tactile feedback and key-press sounds were set to "high." With the E75, the physical QWERTY keyboard and ITU keypad were used with loud key-press sounds. The default audio and tactile feedback of these devices were used. Predictive text entry was not allowed.

Blockings

To block the vision, a piece of cardboard was placed under the subject's chin. The subject was still able to maintain a natural sitting position in the chair. The keyboard of the computer was covered with cardboard so that the subject could not check the QWERTY layout from it.

Hearing was blocked by turning the key-press sounds off, and hearing protectors (Peltor Optime H520) were employed to eliminate the feedback of the natural mechanical sounds.

We used a thin layer of plastic on the keypads that effectively blocked most tactile feedback; in the case of the 5800, also the tactile feedback feature was turned off. We printed the keypads' layout and placed the letter-labeled printout on the plastic layer (see Figure 3). One caveat is that the outer edges of the device could still be felt, although individual keys could not. Another was that the 5800 gives visual feedback on the keyboard: a key flashes when it is pressed. The plastic layer occluded this feedback.

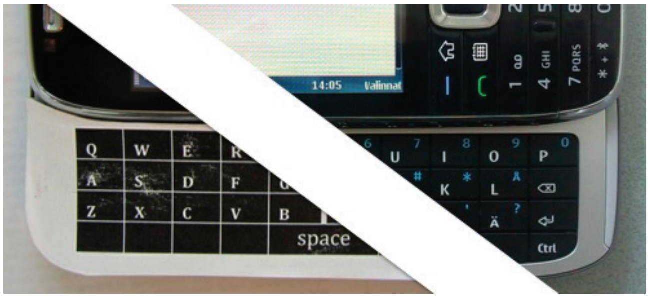

Figure 3. A thin layer of plastic on the keyboard (left) was used to block feedback from the button edges and key releases of the original keyboard (right). Keymaps were printed on top of this layer (physical-QWERTY input shown here).

Design

The experimental design was an eight-by-three within-subjects design with blocking combinations as the first factor and input interface as the second. In total, there were eight modality conditions: *Ø*, *a*, *t*, *v*, *at*, *av*, *tv*, and *atv*, with two trials performed in each. Every subject thus completed 48 trials. The order of the two factors was counterbalanced, by reversing for blockings and by rotating for interfaces. With our decision to keep the none-blocked and all-blocked conditions at the end of a trial, the design yielded a minimum of *n*=12. In the end, the placement of the all-blocked condition at the end was a slight mistake: Despite our attempts to minimize learning effects, users' performance improved during the trials and the scores were higher than if the condition had been counterbalanced.

The experimental design and sample size were planned such that a small-to-medium effect size of 0.4 could be reliably captured for MFI and for D-values, with the aim of a power of 95%. However, for the individual cells of the design, effect sizes would be lower—“medium,” or about 0.6 to 0.7—because the comparisons would be based on fewer samples per user.

Procedure

The subjects were trained to use each keypad via a three-task training set. They were instructed to write the words as correctly as possible and to separate words and sentences with space characters. Correction (backspacing) was forbidden, to minimize variance due to strategic differences and to ensure comparability of blocking conditions.

Before every blocking combination, the subject had a chance to practice with the blocking. When the subject was ready, the moderator made the set of sentences visible. After 30 seconds, a red indicator flashed to mark the end of the time.

All trials were videotaped with a recorder placed on a 1.5 m tripod one meter to the right of the subject.

Measurement

As the performance variable we chose 80% correct words transcribed in 30 seconds, with the idea that 80% correct text messages would still be mostly understandable for the receiver. Moreover, because of blocking of feedback (on the display) in vision-blocking conditions, 100% correct was not realistic. Similar to the Levenshtein metric [20], the figure was calculated by subtracting the number of letter deletions, insertions, and reversals from each word’s length and dividing the result by the presented word’s length.

Table 2. Performance scores in the eight blocking conditions for the three input interfaces.

Condition	Touchpad–QWERTY	Physical–ITU12	Physical–QWERTY
avt	6.79	4.13	7.83
av	4.92	2.50	5.88
vt	6.04	4.29	7.50
at	0.21	1.79	6.04
a	0.21	0.83	0.17
t	0.29	2.00	0.33
v	3.71	2.71	3.71
Ø (none)	0.00	1.21	0.29

Results

In this section of the paper, we utilize the 80%-correct-words-over-30-seconds measure described above. Analogous results were obtained with alternative variables such as 100% correct words per unit time.

Absolute performance

In absolute performance, the physical–QWERTY interface was best, the touchpad–QWERTY one was second best, and ITU12 was worst—with 40% lower performance than the best. The first row in Table 2 indicates the none-blocked situation, the baseline for scaling the scores for the index.

Multimodal flexibility index

An RM-ANOVA was run for MFI, showing a significant effect of the interface, $F(2,22)=5.4$, $p < .05$. Figure 4 shows the situation. With the data subjected to a planned comparison (Scheffe’s test), ITU12 was shown to be distinct from touchpad–QWERTY (p=.016) but not from physical–QWERTY (p=.09), and the two QWERTY interfaces were not statistically different from each other (p=.70).

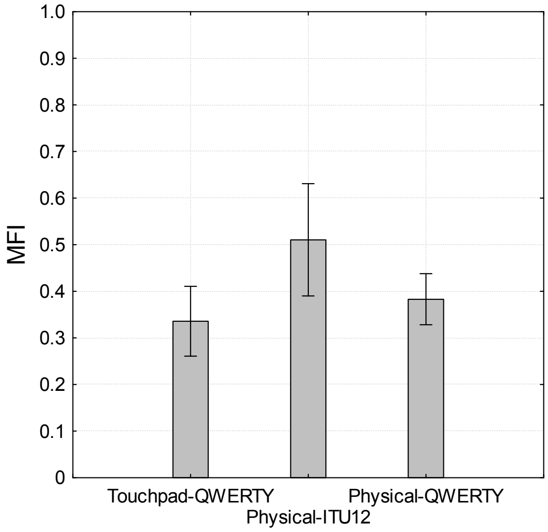

Figure 4. Multimodal flexibility indices for three mobile input interfaces. MFI is calculated according to Definition 1. Vertical bars denote 95% confidence intervals (CIs).

D-values

The D-values obtained are presented in Figure 5. An ANOVA was run with interface and modality as factors, showing a significant interaction effect, $F(4,44)=6.8$, $p < .01$. All three interfaces were vision-dependent (i.e., the D-values were over .50), but the ITU12 interface showed this effect the least. A probable explanation is that the fingers get lost in the middle parts of the QWERTY keyboard when it cannot be seen, while the ITU12 layout is so simple that one can always infer the buttons on which the fingers are resting.

Audition in general was not influential, and adding the other modalities did not change performance. Curiously, D_a was negative for ITU12, which indicates that hearing auditory feedback *decreased* performance. Some users were startled by auditory feedback (beeps) that they were not used to. It may also be that the feedback latency is not optimal. Regardless, the effect was small.

Bimodality analysis

Since D_a was very small for all three interfaces, which indicates that performance does not significantly depend on it, we were left with two modalities to examine: tactition (t) and vision (v). Averaging over the three interfaces, we get s_t=0.88, s_v=3.82, s_{tv}=5.94. Hence, $s_{tv} > s_t+s_v$, which constitutes a case of the two modalities being *synergistic* (see also [30]). This makes sense, because the two modalities aid each other in the task of localizing the position of fingers on the keys and together enable a better "micro-strategy" [11]: vision can be used to monitor feedback on the display and release the fingers to move toward the next buttons without the need to wait for button release.

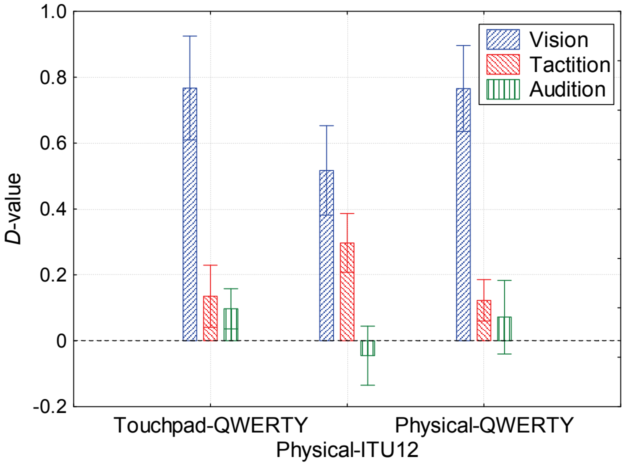

Figure 5. D_m indicates the dependence of performance on m. It is calculated as the average decrease in performance when that modality is blocked (Definition 2). Vertical bars are 95% CIs.

Individual differences

Because all subjects save one were experienced with ITU12 and physical-QWERTY, prior experience was not a predictive factor for these two interfaces. By contrast, only five (of 12) had prior touchpad-QWERTY experience, and these users were, on average, more flexible with this interface than others were (mean MFI of 0.4 as compared with 0.3). However, this difference was not statistically reliable (p=0.17). Frequency of phone use (self-reported) was not a reliable predictor either. One curious finding was that the only heavy (> 100 SMS / month) user had the best mean MFI with all three interfaces (0.57) and a very high MFI for ITU12 (0.74), which he used as his main interface.

DISCUSSION AND RECOMMENDATIONS

In human factors research, a legend was passed on from one paper to another for almost two decades. According to the legend, 90% of all information used in driving is visual [32]. Later on, making such estimations was criticized as meaningless and operationally impossible [35]. Have we attempted something similar here, to propose quantification for the importance of a modality? The answer is both "yes" and "no." Had we done an MFI study of driving, we would have most likely found a very high D_v, but the interpretation would be very different. The correct interpretation of D_v=.90 would be this: "Over all conditions from which vision is removed, the average cost is a 90% performance decrement." "The legend" was essentially a statement about a modality's intake of information in proportion to other modalities, while MFI and its derivatives have a precise meaning as performance changes induced by a modality being withdrawn or blocked from the main task.

The definition of MFI captures a wide range of phenomena that characterize multimodal flexibility beyond a single number, however. For example, the text input study shows an interesting crossover: While ITU12 was the worst interface in terms of absolute performance, it was less hampered by the blockings we administered, within its performance range. Its absolute performance was better than other interfaces' when vision was blocked. The source of this advantage was the use of tactition to compensate for lack of vision. Consequently, ITU12's performance was more dependent on tactition than the others'. Tactition and vision were found to operate synergistically, boosting performance beyond the sum of the two single-modality conditions. By contrast, auditory feedback was not successful: It did not increase performance when vision or tactition was blocked. However, the interfaces may not have optimally used auditory feedback.

A fair criticism can be raised that blocking an exteroceptive sense is crude. First, blocking a sense does not reflect the requirements of typical secondary tasks and environmental conditions, not all secondary tasks require full and uninterrupted allocation. For example, one glance at the speedometer takes under 1.0 s and is enough to inform of driving speed. Second, blocking in our study was an all-or-none business and did not leave room for the strategies users apply in allocating their modalities. For example, switching back and forth among channels (time-sharing) is not possible and we would not be able to observe still finer "micro-strategies" [5, 11]. An exception is tactition, which we blocked in a graded fashion—the thin plastic layer did not prevent feeling the edges of the device. MFI can accommodate finer-grained manipulations (see Table 1 for ideas), but the effects of "bandwidth" allowed by a blocking is a topic for future study. Moreover, future work should address how the results obtained generalize to real-world tasks where modalities can be allocated in whole or in various combinations. A third and related problem is that blocking may not reflect the real bottlenecks of multitasking, such as interference between tasks utilizing the same processing resources or codes [29, 36]. These problems are real but can be addressed by two means: knowing when *not* to use MFI and choosing blockings carefully. The method best suits the analysis of those modalities that are heavily competed for and can be allocated away or blocked for long periods of uninterrupted time. One must accept that, for example, central interference is not addressed by the method.

It is noteworthy that blocking a sensory transduction channel from interaction restricts the usefulness of the method. Because of this characteristic, the method does not suit the study of "intra-interface" multimodality—that is, the deployment of modalities *within* an interactive task. This is one of the main interests of multimodal interface developers. MFI data tell nothing of whether the modalities are used in a cascading [28] or concurrent [26] fashion in commanding the interface. What the indices do indicate is which sensory modalities are available for something else—an "extra-interface" aspect of multimodality.

These discussions and our experiences from the study are summed up in a list of ten recommendations in Table 3.

Table 3. Ten good practices in application of the method.

1. Use the method when you suspect that a modality is important but do not know *how* important. The indices answer the "how much" question, though they tell little about the *why* and *what*. The method does not suit the study of "focal" modalities, whose blocking would take performance to the floor. Understanding the relationship between possible blockings and the sensory *systems* is a topic for future study.
2. The choice of blockings is critical. One should include in the study only modalities that are effective and competed for or challenged in real-world conditions. No information is gained from studying a modality that can always be allocated or is completely passive. Including ineffective blockings will boost the index and convey a falsely optimistic view.
3. Inspect data for instances wherein the blocking of a modality has *improved* performance. The indices assume that performance *decreases* as a result of blocking. This assumption may not always hold, especially where the UI's support for a modality is so poorly designed that blocking the corresponding sensory modality helps the user to achieve better performance.
4. Inspect the indices in light of absolute performance. A high index can be an artifact of performance being at the floor level, which compresses the variability of performance and thereby increases the index. Also, an exceptionally good/bad mean in one condition may pull the index up or push it down in relation to others.
5. Remember that the indices treat all modalities equally. If blockings is not equally distractive, as in our study, where blocking of tactition was only partial in comparison to vision and audition, comparisons of absolute D-values is not recommended. If there are *a priori* reasons for favoring a modality (for example, it is more critical in real-world use or its blocking differs from others'), scores can be weighted.
6. Understand that the indices are contingent on the particular task and the users' skill levels therein. Inspect individual variation in the indices, for example, by examining such variables as skill, prior experience, and exposure. The use of the indices in different types of tasks—closed-loop tasks (e.g., driving), open-loop tasks, alarms, more complex cognitive activities, etc.—is a question for future research.
7. Avoid comparison of indices obtained in a different task, with different blockings, or with different dependent variables.
8. Mix in additional methods such as think-aloud, interviews, and video analysis to obtain qualitative understanding of the events that underlie the indices.
9. Interpret the indices as indicators of how flexible the user is for allocating modalities elsewhere. Optimal dual-tasking in a concrete situation will be contingent on factors not visible from these indices.
10. Pursue other means for further study of the role of a modality in multitasking.

CONCLUSIONS

Most HCI situations engage more than one of the human sensory modalities. We have presented and empirically investigated a generalization of the modality-blocking methodology in order to quantify an important aspect of multimodal user performance: how dependent the user's performance is on modalities being fully allocated to the task. The indices calculated are useful for gauging 1) the dependence of performance on a modality, 2) the cooperation of two modalities, and 3) the overall flexibility with which users are able to reallocate modalities from the interface to other tasks without compromising performance. The method complements existing methods by providing a precise way of assessing this aspect of multimodality for a given interactive task in a way that allows comparisons of interface solutions. We have presented an example study in mobile text input and discussed the limitations of the method, concluding that this method may be best suited to early-stage evaluations of interface solutions. Future work should address the generalizability of indices to real-world HCI.

ACKNOWLEDGEMENTS

This work was funded by the Tekes project Theseus and by the Emil Aaltonen Foundation. We thank Ville Nurmi, Miikka Miettinen, Tuomo Kujala, Lingyi Ma, Pertti Saariluoma, Tero Jokela, Saija Lemmelä, Kimmo Rantala, Jari Laarni, Eve Hoggan, Poika Isokoski, Miika Silfverberg, Céline Coutrix, Heikki Summala, Johannes Tarkiainen, and Shamsi Iqbal for their help and comments. A manual and a sheet for calculations are available at http://www.hiit.fi/mfi.

REFERENCES

1. Baber, C., and Mellor, B. Using critical path analysis to model multimodal human–computer interaction. *International Journal of Human–Computer Studies 54*, 4 (2001), 613-636.
2. Bertelson, P., and de Gelder, B. The psychology of multimodal perception. In: *Crossmodal Space and Crossmodal Attention* (2004), 141-177.
3. Brewster, S., Lumsden, J., Bell, M., Hall, M., and Tasker, S. Multimodal "eyes-free" interaction techniques for wearable devices. *Proc. CHI2003*, ACM Press (2003), New York, pp. 473-480.
4. Briem, V., and Hedman, L. Behavioural effects of mobile telephone use during simulated driving. *Ergonomics 38*, 12 (1995), 2536-2562.
5. Brumby, D., Howes, A., and Salvucci, D. A cognitive constraint model of dual-task trade-offs in a highly dynamic driving task. *Proc. CHI2007*, ACM Press (2007), New York, pp. 233-242.
6. Burke, J., Prewett, M., Gray, A., Yang, L., Stilson, F., Coovert, M., Elliot, L., and Redden, E. Comparing the effects of visual–auditory and visual–tactile feedback on user performance: A meta-analysis. *Proc. ICMI2006*, ACM Press (2006), New York, pp. 108-117.

7. Cashdan, S. and Zung, B. J. Effect of sensory modality and delay on form recognition. *Journal of Experimental Psychology 86*, 3 (1970), 458.

8. Cohen, M. R., and Nagel, E. *Logic and Scientific Method*. Harcourt Brace Jovanovich (1934), New York.

9. Colquhoun, W. Evaluation of auditory, visual, and dual-mode displays for prolonged sonar monitoring in repeated sessions. *Human Factors 17*, 5 (1975), 425.

10. Driver, J., and Spence, C. Cross-modal links in spatial attention. *Philosophical Transactions: Biological Sciences 353*, 1373 (1998), 1319-1331.

11. Gray, W., and Boehm-Davis, D. Milliseconds matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. *Journal of Experimental Psychology Applied 6*, 4 (2000), 322-335.

12. Hoggan, E., Brewster, S., and Johnston, J. Investigating the effectiveness of tactile feedback for mobile touchscreens. *Proc. CHI2008*, ACM Press (2008), New York, pp. 1573-1582

13. Hoggan, E., Raisamo, R., and Brewster, S. Mapping information to audio and tactile icons. *Proc. ICMI2009*, ACM Press (2009), New York.

14. Isokoski, P., and Linden, T. Effect of foreign language on text transcription performance: Finns writing English. *Proc. NordiCHI2004*, ACM Press (2004), New York, pp. 109-112.

15. Jacko, J., Barnard, L., Kongnakorn, T., Moloney, K., Edwards, P., Emery, V., and Sainfort, F. Isolating the effects of visual impairment: Exploring the effect of AMD on the utility of multimodal feedback. *Proc. CHI2004*, ACM Press (2004), New York, pp. 311-318.

16. Jaimes, A., and Sebe, N. Multimodal human–computer interaction: A survey. *Computer Vision and Image Understanding 108*, 1-2 (2007), 116-134.

17. Kane, S., Wobbrock, J., and Smith, I. Getting off the treadmill: Evaluating walking user interfaces for mobile devices in public spaces. *Proc. MobileHCI*, ACM Press (2008), New York, pp. 109-118.

18. Kieras, D., and Meyer, D. An overview of the EPIC architecture for cognition and performance with application to human–computer interaction. *Human–Computer Interaction 12*, 4 (1997), 391-438.

19. Lee, J., and Spence, C. Assessing the benefits of multimodal feedback on dual-task performance under demanding conditions. *Proc. BritishHCI2008*, British Computer Society (2008), Swindon, UK, 185-192.

20. Levenshtein, V. Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady 10*, 8 (1966), 707-710.

21. Liu, Y. Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems. *Ergonomics 44*, 4 (2001), 425-442.

22. Lyons, K., Starner, T., Plaisted, D., Fusia, J., Lyons, A., Drew, A., and Looney, E. Twiddler typing: One-handed chording text entry for mobile phones. *Proc. CHI2004*, ACM Press (2004), New York, pp. 671-678.

23. MacKenzie, I. Motor behaviour models for human– computer interaction. In J. Carroll (ed.), *HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science*. Morgan Kaufmann (2003), San Francisco, 27-54.

24. MacKenzie, I., and Soukoreff, R. Phrase sets for evaluating text entry techniques. *Proc. CHI2003*, ACM Press (2003), New York, pp. 754-755.

25. Mizobuchi, S., Chignell, M., and Newton, D. Mobile text entry: Relationship between walking speed and text input task difficulty. *Proc. MobileHCI2005*, ACM Press (2005), New York, pp. 122-128.

26. Nigay, L., and Coutaz, J. A design space for multimodal systems: Concurrent processing and data fusion. *Proc. CHI1993*, ACM Press (1993), New York, 172-178.

27. Oulasvirta, A., Tamminen, S., Roto, V., and Kuorelahti, J. Interaction in 4-second bursts: The fragmented nature of attentional resources in mobile HCI. *Proc. CHI2005*, ACM Press (2005), New York, pp. 919-928.

28. Oviatt, S. Multimodal interfaces: The human–computer interaction. In A. Sears, J. Jacko (eds.), *Human-Computer Interaction Handbook*, Lawrence Erlbaum (2003), New York, 286-304.

29. Pashler, H. *The Psychology of Attention*. MIT Press (1999), Boston.

30. Perakakis, M., and Potamianos, A. Multimodal system evaluation using modality efficiency and synergy metrics. *Proc. ICMI2008*, ACM Press (2008), New York, pp. 9-16.

31. Posner, M. *Chronometric Explorations of Mind*. Erlbaum (1978), Hillsdale, NJ.

32. Rockwell, T. Skills, judgment and information acquisition in driving. *Human Factors in Highway Traffic Safety Research*, Wiley (1972), New York, 133-164.

33. Salvucci, D. A multitasking general executive for compound continuous tasks. *Cognitive Science 29*, 3 (2005), 457-492.

34. Senders, J. W., Kristofferson, A. B., Levison, W. H., Dietrich, C. W., and Ward, J. L. The attentional demand of automobile driving. Highway Research Record #195. National Academy of Sciences, Transportation Research Board (1967), Washington, DC, 15-33.

35. Sivak, M. The information that drivers use: Is it indeed 90% visual? *Perception, 25*, 9 (1996), 1081-1090.

36. Wickens, C. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science 3*, 2 (2002), 159-177.