Book:

Handbook of Multimodal-Multisensor Interfaces

(Multimodal Processing of Attention, Cognition, and Expertise**)**

Editor: Sharon Oviatt

1 Multimodal Behavioral and Physiological Signals as Indicators of Cognitive Load

Jianlong Zhou, Kun Yu, Fang Chen, Yang Wang, Syed Z. Arshad
Affiliation: Data61, CSIRO, Australia
Email: {jianlong.zhou, kun.yu, fang.chen, yang.wang, syed.arshad}@data61.csiro.au

1.1 Introduction

Cognitive load refers to the load on working memory as experienced by subjects when conducting cognitive tasks. It is a critical factor affecting human behaviour in modern complex mission-critical domains such as aviation, military command and control [Staal 2004]. Extreme cognitive load causes stress, results in human errors, and limits a human's ability to conduct a given task, especially when the task demands exceed the capacity of the human working memory. In contrast, a manageable level of cognitive load will have a lower implication on the subjective performance. As a consequence, the capability to measure a human's cognitive load can provide insights for system adaptation to maximize user's performance - thereby easing the cognitive demand (from user) and avoiding stress, frustration and ultimately errors.

A great variety of cognitive load measuring techniques, from simple questionnaires to complex functional brain imaging [Whelan 2007] , have been developed to study cognitive load. Following are four primary methods comprising the current state of the art [Chen et al. 2016]:

- **Subjective (self-report) measures**, where users rank their experienced level of load on single or multiple rating scales [Gopher and Braune 1984];
- **Performance measures**, such as task completion time, speed or correctness, critical errors and false starts [Gawron 2000; O'Donnell and Eggemeier 1986; Paas et al. 2008], as well as performance on a secondary tasks [Chandler and Sweller 1991];
- **Physiological measures**, such as galvanic skin response, and heart rate [Delis et al. 2001; Ikehara and Crosby 2005];
- **Behavioral measures**, which observe feature patterns of interactive behavior, such as linguistic or dialogue patterns [Berthold and Jameson 1999], and even text input events and mouse-click events [Arshad et al. 2013].

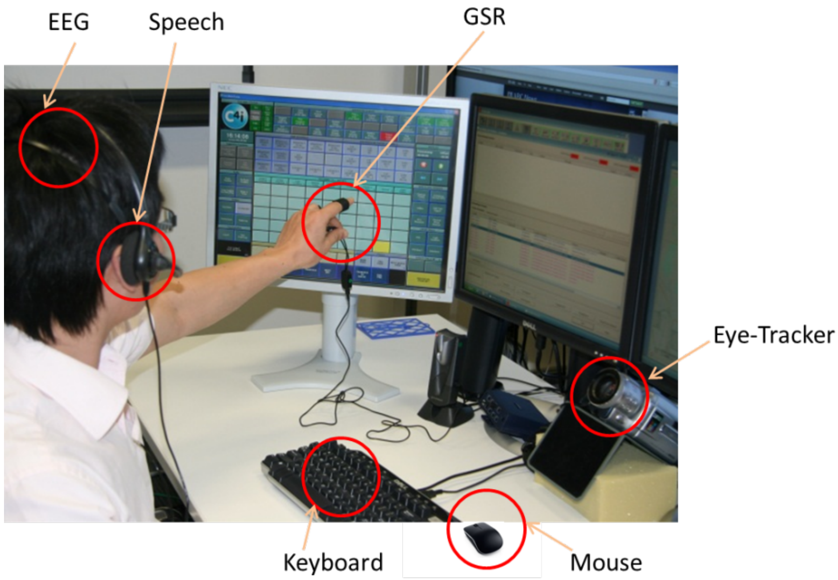


Figure 1-1. Example modalities used for cognitive load measurement.

More specifically, with the advance of modern sensor technologies, various physiological and behavioral measures have been developed for the assessment of cognitive load. Popular methods for cognitive load measurement utilize signals from brain waves, eye activity, respiration, heart rate, skin conductance, speech, etc. These approaches have attracted increasing attention since several of them can provide detailed evaluation of cognitive state in a non-intrusive way. Although various studies exhibit the effects of cognitive load on physiological measures, no single physiological measure is sufficient to comprehensively characterize the cognitive load [Wierwille 1979], especially in the case of multidimensional task and/or dynamic circumstances. On the other hand, physiological measures may be affected by many other factors, such as engagement, fatigue, stress, and environment. Therefore, *cognitive load measurement* based on multiple modalities has been adopted to improve the robustness and accuracy [Chen et al. 2016]. As shown in Figure 1-1, signals from various sensors and input devices can be collected during cognitive tasks.

This chapter evaluates cognitive load from the perspective of human responses, which is a *data-driven* approach for cognitive load measurement. Human responses can be characterised by various behavioral and physiological signals which are recorded and analysed. Taking the speech modality as an example, it is a natural form of communication between human beings. Besides linguistic interpretation, speech also conveys information such as speaker identity and mental state related information, e.g. cognitive load [Yin et al. 2008]. As shown in Figure 1-2, speech is an acoustic signal, generated by the airflow from the lungs considered to be the voice source which then passes through to the pharynx and the oral and nasal cavities, collectively known as the vocal tract filter. The features of the voice source and the vocal tract filter vary based on the content of the utterance to be pronounced as well as the mental state of the speaker [Yin et al. 2008].

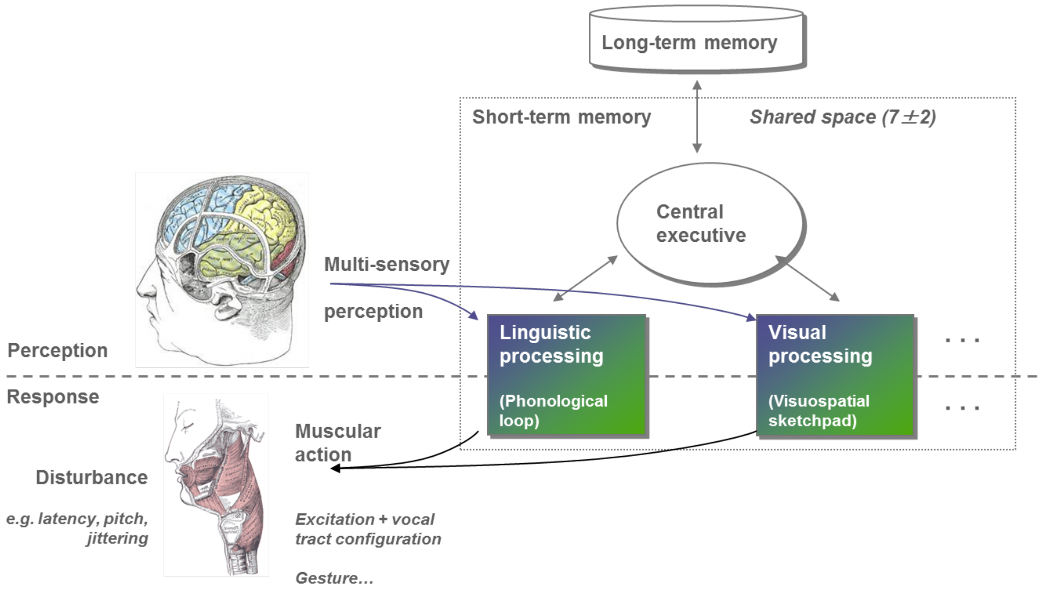


Figure 1-2. Speech generation and its relationship with cognitive load.

In Section 1.2, we firstly investigate state-of-the-art theories and approaches for cognitive load measurement as well as applications of cognitive load measurement. Behavioral measures of cognitive load are presented in Section 1.3, which include pen features, speech, and linguistic features. In Section 1.4, various physiological measures are introduced for cognitive load measurement. We focus on the pupil dilation and *Galvanic Skin Response (GSR)* for measuring cognitive load in this chapter. Section 1.5 presents a framework of dynamic workload adjustments in a feedback loop to help users maximize their capacity for productive work. We finally give conclusions in Section 1.6.

For a definition of terms introduced in this chapter see the Glossary table.

Glossary

Accumulative GSR refers to the summation of GSR values over the task time. If the GSR is considered as a continuous signal over time, the accumulative GSR is the integration of GSR values over the task time.

Cognitive load is a multidimensional construct that refers to the momentary working memory load experienced by a person while performing a cognitive task. It can be increased by a wide variety of factors, including the difficulty of a task, the materials or tools used (e.g., computer interface), the situational context (e.g., distracting versus quiet setting), the social context (e.g., working individually, versus jointly with a group), a person's expertise in the domain, a person's physiological stress level, and so forth. Cognitive Load Theory describes cognitive load as having three components—intrinsic load, extraneous load, and germane load [Sweller et al. 2011]. For a detailed discussion of the dynamic and often non-linear interplay between cognitive load and domain expertise, including how cognitive load can either expand or minimize the performance gap between low- and high-performing students, see [Oviatt 2013].

Intrinsic load is the inherent difficulty level and related working memory load associated with the material being processed during a user's primary task.

Extraneous load refers to the level of working memory load that a person experiences due to the properties of materials or computer interfaces they are using [Oviatt 2017].

Germane load refers to the level of a person's effort and activity compatible with mastering new domain content during learning. It pertains to the cognitive resources dedicated to constructing new schema in long-term memory.

Cognitive load measurement refers to the methods to quantitatively discriminate the different levels of cognitive load experienced by the user. Usually the cognitive load is induced with varied task difficulty (i.e. the extraneous load is manipulated), and the methods to discriminate cognitive load include subjective methods, performance-based methods, physiological methods and behavioral methods.

GSR refers to galvanic skin response which is a measure of the conductivity of human skin, and can provide an indication of changes in the human sympathetic nervous system during the cognitive task time.

Multimodal fusion is the technology of synthesizing multiple channels of signals or features to achieve higher precision in cognitive load measurement, or better reliability via overcoming the limitations of individual signal or interaction modalities. The fusion can be done at different stages in cognitive load measurement: mid-fusion and late-fusion. Mid-fusion refers to the fusion of features extracted from multimodalities before cognitive load classifications, while late-fusion is the fusion of classification scores from single modality decisions.

1.2 State-of-The-Art

Recent years have witnessed booming growth in the research of cognitive load understanding and measurement [Chen et al. 2012]. Extensive research has been conducted on cognitive load theory, cognitive load measurement, applications of cognitive load theory, and the factors that may affect cognitive load.

It is well-established that the two main limitations of working memory resources exist, i.e. its capacity and duration [Baddeley 1992]. According to Baddeley's model [Baddeley 2012], working memory has separate processors for visual and verbal information. Only a limited number of items, or "chunks", can be held in working memory at the same time and only for limited duration [Cowan 2001]. These limitations are never more evident than when users undertake complex tasks, or when in the process of learning – extremely high demands are placed on working memory. The construct of cognitive load refers to the working memory demand induced by a complex task in a particular instance where novel information or novel processing is required [Sweller et al. 1998]. It should be noted that different people, although facing the same task, may experience different cognitive load. This inter-personal discrepancy could be attributed to a number of reasons, for example, level of domain expertise or prior knowledge, interface familiarity, age, or any mental or physical impediments. A task resulting in high cognitive load for one user may not necessarily have the same implication for another.

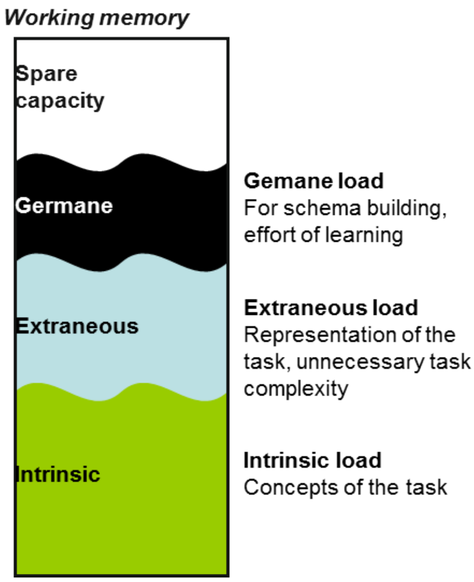


Figure 1-3. Three types of cognitive load.

The cognitive load construct comprises at least three separate load sources: *intrinsic load, extraneous load,* and *germane load* [Sweller et al. 1998; Paas et al. 2003] (see Figure 1-3). Extraneous cognitive load refers to the level of working memory load that a person experiences due to the properties of materials or computer interfaces he/she is using. High levels of extraneous cognitive load can undermine a user's primary task performance. Extraneous load is distinguished from (1) intrinsic cognitive load, or the inherent difficulty level and related working memory load associated with a user's primary task, and (2) germane cognitive load, or the level of a user's effort and activity compatible with mastering new domain content, which may either be supported or undermined by interface design, e.g., due to inappropriate automation. [Sweller et al. 1998; Paas et al. 2003].

The ability to determine exactly when a user is being cognitively loaded beyond a level that they are able to manage could enable the system to adapt its interaction strategy intelligently. For example, the system could attempt to reduce the cognitive load experienced by the human – particularly in terms of extraneous load – such that optimal performance is facilitated. A number of methods have been used, both in Human-Computer Interaction (HCI) and other domains, to estimate the level of cognitive load experienced. Generally, amongst the four categories of cognitive load measurement methods mentioned above, subjective rating has been used as ground truths in cognitive load experiments, but its usage is quite limited in that it can only be conducted after cognitive tasks. Performance measures offer a good temporal relationship to task demands, but are generally insensitive to variations in load until it approaches or exceeds cognitive capacities. Whereas, behavioral and physiological measures hold promise in providing a more direct and immediate method to assess cognitive load. However, establishing a direct relationship between such measures and user's cognitive state can be a challenging task.

1.2.1 Subjective Measures

Traditionally, the most consistent results for cognitive load measurement have been achieved through subjective measures [O'Donnell and Eggemeier 1986]. These measures reflect each user's perception of cognitive load experienced by means of introspection. The user is required to perform a self-assessment of their mental demand by answering a set of assessment questions immediately after the task.

There are two types of scales of subjective ratings:

- **Unidimensional scales**, which measure overall cognitive load like subjective cognitive load measurement scale [Paas and Merriënboer 1994].
- **Multidimensional scales**, which focus on the different components of load.

According to cognitive load theory (CLT), unidimensional scales are criticized because they do not do justice to the multifactorial nature of cognitive load. It is also nevertheless recognized that a single item of difficulty is a good indicator of overall cognitive load [Debue and van de Leemput 2014]. On the other hand, a multidimensional scale such as the NASA Task Load Index (NASA-TLX) [Hart and Staveland 1988] gives a broader evaluation of cognitive load. It is based on six dimensions (performance, mental effort, frustration, task demand, physical demand, temporal demand). Cierniak et al. [Cierniak et al. 2009] proposed a three-item scale which includes a difficulty rating for the learning content, difficulty of the material and concentration during learning. As a result, it assesses intrinsic, extraneous, and germane load respectively. More recently, Leppink et al. [Leppink et al. 2013] have developed a ten-item subjective cognitive load scale with students attending lectures on statistics. Despite these promising approaches, there are some limitations inherent to self-reported measures. For instance, subjective scales are often administered after the learning task and, in this way, are not easily able to capture variations in load over time [Paas and Merriënboer 1994]. Such an approach is also impractical in realistic situations because the questionnaires not only interrupt task flow but also add extra tasks to an already potentially overloaded user [Chen et al. 2012].

NASA-TLX is available both online and as mobile APPs on both Android and iOS:

- https://humansystems.arc.nasa.gov/groups/tlx/
- https://itunes.apple.com/us/app/nasa-tlx/id1168110608
- https://play.google.com/store/apps/details?id=org.texoft.nasa_tlx&hl=en
- http://www.nasatlx.com/

Figure 1-4 shows the example of a NASA-TLX APP on iOS.

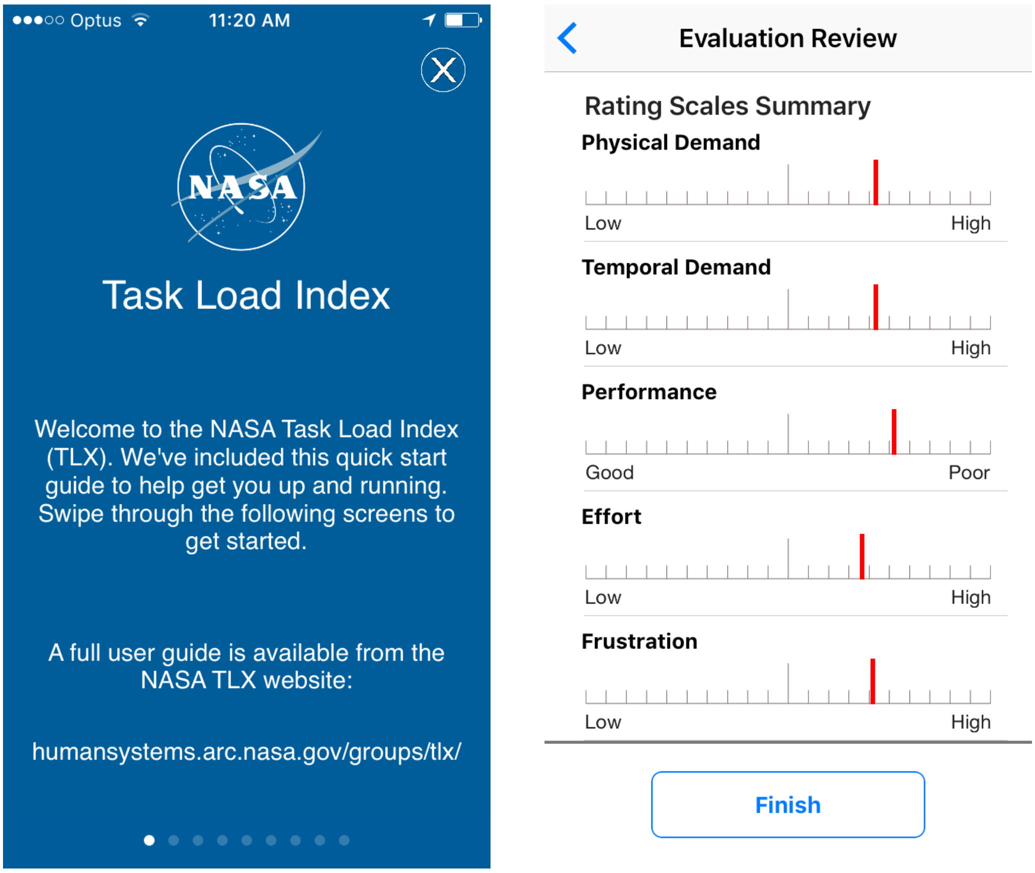


Figure 1-4. A NASA-TLX APP on iOS (https://itunes.apple.com/us/app/nasa-tlx/id1168110608).

1.2.2 Performance Measures

Performance-based approaches provide measures that can reflect variations encountered during the task. Typical performance measures include task accuracy, time used to complete the task and so on. In real world tasks, performance measures from the primary task can be extremely difficult to calculate on the fly, if at all. The basic assumption for performance measures is that learning is hindered when working memory capacity is overloaded, and therefore a drop in performance will be the result of the overall high cognitive load [Paas and Merriënboer 1994]. One of the most widely used techniques is the dual-task paradigm, in which the performance is evaluated in a secondary task performed in parallel to assess the cognitive load devoted to the main task. The hypothetical relationship between performance and workload as discussed by O'Donnell and Eggemeier [O'Donnell and Eggemeier 1986], is composed of three regions, Low, Medium and High. In the "Medium" region, both primary and secondary task performance measures can be used to reflect workload increase as performance decreases. Dual-task approaches have been incorporated in several studies to measure subjects' performance in controlled conditions [O'Donnell and Eggemeier 1986]. While secondary task performance can provide a measure of remaining resources not being used by the primary task [Kerr 1973; Marcus et al. 1996], it is not feasible for humans to complete dual tasks "in the wild", and hence these cannot be adopted for widespread use. Performance measures tend to remain stable as load increases in the "Medium" region, particularly when the human exerts a greater amount of cognitive load, as noted by [O'Donnell and Eggemeier 1986].

Performance measures – which are often defined as measures that reflect the accuracy and correctness of a user's response and are directly relevant to the outcome of the task – are often calculated after the fact, if they can be assessed objectively at all. In some complex domains, measures based on performance outcomes are impossible to access in real time such that the system would be able to act on the information in a timely manner. For example, the spontaneous nature of crisis management and other control room situations means the user's performance in this sense is very difficult to rate, even during debriefing, and unique to almost every situation. The actions taken can vary widely from human to human, both in order and content, while still being equally effective in achieving the task goals and solving the problem to an adequate level of performance. In some cases, performance cannot be calculated automatically at all [Chen et al. 2012].

1.2.3 Factors that Affect Cognitive Load Measurement

As mentioned in the previous section, Leppink et al. [Leppink et al. 2013] presented a ten-item instrument to measure three types of cognitive load. Task complexity and the subject's prior knowledge determine the intrinsic load, instructional features that are not beneficial for task completion contribute to extraneous load, and instructional features that are beneficial for task completion contribute to germane load. Efforts have been made to measure different types of cognitive load respectively. In [Leppink et al. 2013], the data were collected during four lectures and in a randomized statistics experiment with 0-10 rating scales for CLM, via a survey involving ten questions aiming to evaluate three types of cognitive load. The work was further extended in [Leppink et al. 2013] by adapting the survey instrument to the domain of learning languages.

However, different types of load are not independent from each other. Morrison et al. also [Morrison et al. 2014] adapted the Cognitive Load Component Survey in [Leppink et al. 2013] in an introductory computer science context. Debue and Leemput [Debue and van de Leemput 2014] examined the three types of cognitive load via manipulating the amount of multimedia elements of an online newspaper. In examining the relationship between the performance and cognitive load factors, they showed the expected opposite relationships between germane and extraneous load, which means that when contents are not presented clearly to the users, they contribute less to their learning. Debue and Leemput [Debue and van de Leemput 2014] also showed a positive association between germane load and cognitive absorption (which is defined as "a state of deep involvement with software" [Agarwal and Karahanna 2000]) and an association between intrinsic and germane load.

Researchers also investigated cognitive load differences between age and gender groups. Ferreira et al. [Ferreira et al. 2014] assessed cognitive load based on psycho-physiological measurements for younger and older adults. Two different types of tasks were used: the Pursuit Test (PT), in which different curves are entangled and subjects were required to determine where a curve begins and ends, and the Scattered X's test (SX), in which the letter "X" was displayed together with other letters on the screen, in which subjects were required to locate all the X's in a display. These tests measure perceptual speed and visio-spatial cognitive processing capabilities identified from the fields of psychology and cognitive science. Participants' psycho-physiological responses were recorded with four sensor devices: GSR, ECG (records heart rate and breathing rate (BR)), EEG, and an armband (records heat flux – rate of heat transfer on the skin). It was found that the EEG signal was a better predictor and the BR measurement less important for the older than for the younger participants. Heart rate and breathing rate signals from raw ECG signal were also fairly well represented among the most common features for both age groups, whereas GSR features were seldom used for the younger participants and never for the older participants. Verrel et al. [Verrel et al. 2009] found that different age groups showed different gait patterns with high cognitive load during treadmill walking. For example, when cognitive load was increased, gait patterns became more regular in those 20–30 years old, less regular in those 70–80 years old, and showed no significant effects in those 60–70 years old.

Furthermore, information presentation methods may also affect the difference of cognitive load. Dindar et al. [Dindar et al. 2014] developed test questions either with static graphics or with animated graphics accompanied with text to measure students' cognitive load during the learning process. Students' response time, response accuracy, self-reported ratings on cognitive load and secondary task performance were used to measure their cognitive load. It was found that animating graphics increased the response time and secondary task scores (higher secondary task scores relate to higher cognitive load levels) of the students but did not have any significant effect on their test success. Self-ratings and response accuracy were also found more sensitive to intrinsic cognitive load, whereas response time and secondary task measures were found to be more sensitive to extraneous cognitive load.

1.2.4 Applications of Cognitive Load Measurement

Monitoring the mental state of users can profoundly improve interactions and the user experience in today's autonomous systems. Various practical aspects of a person's life including user interface design, education and training, transportation (road, rail, sea or air), emergency management, etc. would dramatically benefit from objective, robust, accurate, real-time, unobtrusive detection of cognitive load.

Sweller [Sweller 1994] demonstrated that ineffective user interface designs may interfere with information acquisition by increasing the associated cognitive load. Back and Oppenheim [Back and Oppenheim 2001] found that users would prefer an interface design requiring a relatively low cognitive load that at the same time, can reach high user satisfaction. Shi et al. [Shi et al. 2010] proposed a Cognition-Adaptive Multimodal Interface (CAMI), for a large metropolitan traffic incident and emergency management system. CAMI aims to simplify the operations and provide users with a new user interface that can sense user's cognitive load and provide cognitive support adaptively. CAMI involves the following components: a multimodal interface which facilitates user's input to the system and detection of user's cognitive state; an input and context analyzer which analyzes user's intention and cognitive load level; an output and support part which provides user with information display and tools based on user needs and context (see Figure 1-5).

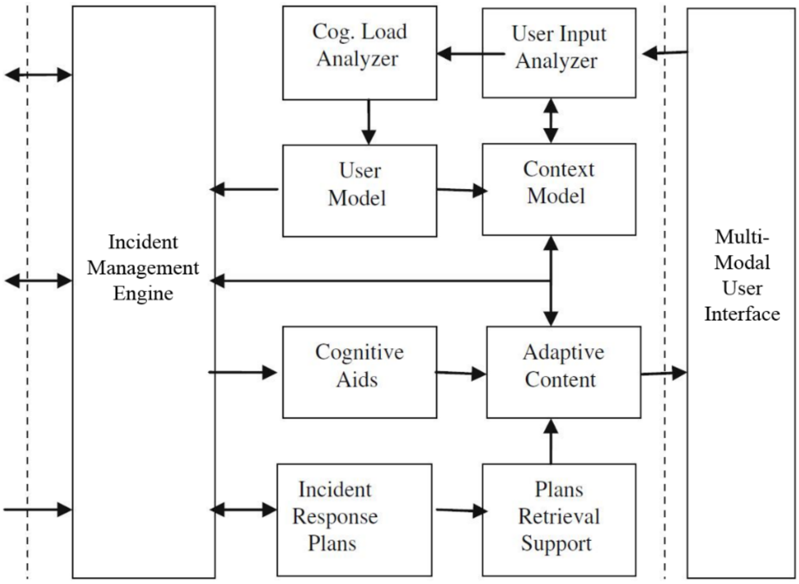


Figure 1-5. CAMI consists of a multimodal user interface and a decision-supporting middleware (between the two dotted lines) [Shi et al. 2010].

Emergency management is one of the extremely important domains in which a human's cognitive fatigue or mental overload can result in irreversible consequences. Khawaja et al. [Khawaja et al. 2014] investigated cognitive load of humans when they were conducting bushfire emergency management tasks via a simulated study. Linguistic features during bushfire management tasks were analyzed and it was found that during the emergency management participants showed significant differences in linguistic patterns between low and high cognitive load tasks, where an increase in the totally number of words spoken and the number of words used per sentence were observed in high cognitive load conditions [Khawaja et al. 2014]. By monitoring human's cognitive load during emergency management implicitly, human's activity can be adaptively modulated in order to improve the working efficiency. Other emergency management domains such as traffic management and call centers can also benefit from monitoring a human's cognitive state and therefore avoiding the mistakes caused by their mental overload.

Driver distraction via secondary in-vehicle activities is increasingly recognized as a significant source of injuries and fatalities on the roadway [Strayer et al. 2013]. Extensive research has been carried out in this area. For example, Engstroem and colleagues [Engström et al. 2005] examined the effects of visual cues and cognitive load on driving performance and driver state during simulated and realistic motorway driving. In their study, cognitive load resulted in decreased lane keeping performance and increased gaze concentration toward the road center. In another study, Crundall and Underwood [Crundall and Underwood 1998] investigated the effects of cognitive load on the driving performance of expert and novice drivers on different types of roads. They found that experienced drivers chose visual strategies according to the complexity of the road but the strategies of novices were not flexible enough to respond properly to the variations of cognitive demand provided by each road. Besides car driving, cognitive load measurement holds promise as being a useful tool in monitoring the mental states of cockpit crews, air-craft maintenance teams, and military teams where performance is paramount, and especially for pilots, where the accuracy and speed of communication and operation is critical to safety and where errors may result in the loss of life [Cannon-Bowers et al. 1993].

The limited capacity of working memory based on cognitive load theory (CLT) is a widely recognized determinant of human learning. Hessler and Henderson [Hessler and Henderson 2013] applied CLT to nursing education. Via monitoring the cognitive load and adjusting the curricula information accordingly, their work suggest that it is feasible to improve the learning outcome. Cognitive load can also be manipulated by tailoring instructional design to levels of a learner's prior knowledge [Kuldas et al. 2014]. In this context, the instructional design is not only aimed at controlling the cognitive load, but also at stimulating learners to use their available cognitive capacity for better learning. In this line, Sweller [Sweller et al. 2011] also suggested that encouraging learners to exert more cognitive effort is a feasible means to enhance the construction of schemas.

Cognitive load measurement can be used in many other applications. For example, recent research has shown that cognitive load has an effect on gait [Martin and Bajcsy 2011], affecting the walking speed, which is especially noticeable in human with neurodegenerative disorders such as Alzheimer's disease vascular dementia, mixed dementia [Allali et al. 2008] or Parkinson's disease [Yogev et al. 2007].

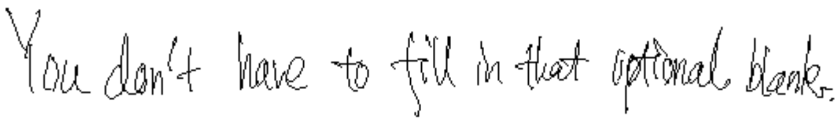
Cognitive load measurement also plays significant roles in improving information retrieval system design [Na 2012]. In examining user interface designs of information retrieval systems, Hu et al. [Hu et al. 1999] utilized cognitive load as a measurement of the information seeking and processing effort. Specifically, they found that via reducing cognitive load, information searchers' satisfaction tended to be improved. Additionally, cognitive load measurement has also been applied in simultaneous interpreting tasks to improve an interpreter's performance by monitoring cognitive load in real time [Seeber 2013].

1.3 Behavioral Measures for Cognitive Load

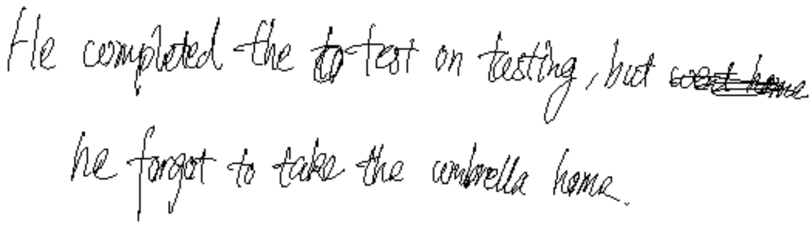
Cognitive load is considered as a key factor affecting the behaviours of people. High cognitive load results in decreased control in human motions, which have been verified via the studies of human hand movements [Yu et al. 2011] and walking gestures [Verrel et al. 2009] when challenging cognitive tasks were conducted. As a consequence, behavioral features, if properly extracted, are expected to be able to reflect cognitive load variations, and thus measure cognitive load.

Behaviours generally refer to the way people physically conduct different tasks, and are mostly concerned with motions of the human body. Examples of behavioral phenomena addressed in this chapter include hand movements when writing, typing on a keyboard or using a mouse, head position changes to examine the road conditions when driving a car, gait and stride when walking, and lip and vocal cord movement when speaking. Over time and repeat experience, many human behaviors have become automated in the brain such that they no longer require focal attention. Instead, people just focus on the high level goal of a task, for example, to interact with the computer without considering using which finger to tap the next key.

It is generally believed that cognitive load has implications for the way human write. Figure 1-6 shows writing examples in sentence composition tasks under different cognitive load levels from a same subject [Yu et al. 2011]. Many writing features (see Figure 1-7) including writing speed, pen-tip pressure, and the way pen is grasped can be used to characterize the cognitive load of the writer. Recent research has shown that people increase their writing speed when high cognitive load is experienced [Yu et al. 2011]. One possible explanation is that writing offloads information being held in short-term memory into a visible form on paper, which reduces working memory demands and related cognitive load experienced during the task. Under high cognitive load conditions, there exists a high demand to speed up the information offloading process.



(a)



(b)

Figure 1-6. Hand writing examples in sentence composition tasks with given words: (a) low load (given word: optional), and (b) high cognitive load (given words: complete, forget, umbrella) from a same subject.

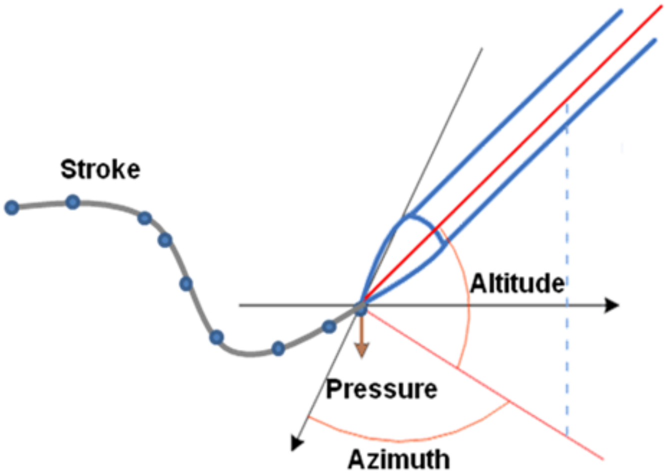


Figure 1-7. Typical features extracted from handwriting.

Similar with handwriting, speech as another means of communication has been examined for its relation to cognitive load [Keraenen et al. 2004; Mueller et al. 2001]. Generally the parameters of the voice source and the vocal tract vary according to the content of the utterance to be pronounced as well as the mental state of the speaker. In Mueller's study, under the dual-task navigation condition subjects speak more slowly and with more disfluencies. In investigations conducted by Phu [Phu 2011], the speaker's breath pattern, speaking rate and filler sounds etc. characterize different aspects of the speech production system and thus contribute to cognitive load information in a different way, as shown in Table 1-1. It should be noted that contrary to Mueller's findings, the speech rate as a feature has dropped when cognitive load is high. This phenomenon may be due to several factors during the investigation: when measuring speech rate Mueller et al. examined articulation rate separately from unfilled pauses. Furthermore, the spoken constructions were controlled to determine whether articulation rate has changed. Otherwise, when under high load, subjects reduce utterance length and complexity, which reduces planning demands and thereby cognitive load.

Table 1-1. Speech cues of three cognitive load levels.

Speech cue \ Cognitive Load	Low Level	Medium Level	High Level
Breath pattern	Soft	Heavy	Heavy
Speech rate (the number of words per unit of time) within utterance	Slow, consistent	Slow, inconsistent	Fast, consistent
Filler sounds e.g. 'uh' and 'ah'	Never	Sometimes	Sometimes
Intonation	Naturally varying	Less varying	Flat

Other features such as linguistic and grammatical features, may also be extracted from user's spoken language for patterns that may be indicative of high cognitive load. Significant increase in the level of spoken disfluencies but decrease in speech articulation rates have been found when users experience high cognitive load, as summarized in [Berthold and Jameson 1999]. Extensions of this work have identified increased usage in first person plurals [Sexton and Helmreich 2000] as indices of cognitive load.

BrainGauge is a commercial product available to assess cognitive load from subject's speech voice in real-time. More information on BrainGauge can be found from:

- https://www.braingauge.com.au/

1.4 Physiological Measures for Cognitive Load

Research has found that changes in various physiological states are closely related to changes in cognitive load [Chen et al. 2012]. One major advantage of physiological measures is the continuous availability of bodily data, allowing cognitive load to be measured at a high temporal rate with improved precision. This research has also shown that different brain regions control different cognitive and mental activities, for instance, the dorsolateral prefrontal cortex (DLPFC) is associated with intrinsic load in cognitive activities [Whelan 2007]. The functional imaging and clinical evidence indicates that a remarkably consistent network of brain regions is involved in cognitive activities [Moll et al. 2005]. Van Gog et al. [Gog et al. 2008] used an approach combining evolutionary biological theory and neuroscience [Ayres and Paas 2009] to explain human's behaviour in observational learning. Because of the close relationship between cognitive load and neural systems, human neurophysiological signals are seen as promising avenues to measure cognitive load based on a number of measurement criteria: sensitivity, diagnosticity, intrusiveness, reliability, and generality of application [Kramer 1991].

The physiological approach for cognitive load measurement is based on the assumption that any changes in the human cognitive functioning are reflected in the human physiology [Kramer 1991]. The measures that have been used in the literature to show some relationship between subjects' mental workload or cognitive load and their physiological signals including:

- Heart rate and heart rate variability [Kennedy and Scholey 2000; Mousavi et al. 1995; Nickel and Nachreiner 2000];
- Brain activity (e.g. changes in oxygenation and blood volume, electrocardiography (ECG), electroencephalography (EEG)) [Brunken et al. 2003; Wilson and Russell 2003];
- Galvanic skin response (GSR) or skin conductance [Jacobs et al. 1994; Shi et al. 2007];
- Eye activity (e.g. blink rate, pupillary dilation, and peak dilation) [Backs and Walrath 1992; Iqbal et al. 2004; Lipp and Neumann 2004; Marshall et al. 2003; Chen and Epps 2013; Wang et al. 2013]. These pupillary responses typically intensify as a function of cognitive load.

Changes in the physiological data occur with the level of stimulation experienced by the person and can represent various levels of mental processing. The data collected from body functions are useful as they are continuous and allow the signal to be measured at a high rate and in fine details. However, physiological measures normally require users to wear cumbersome equipment, e.g. EEG headsets that not only interfere with users' task, but are prohibitive in cost and deployment. Additionally, the large amounts of physiological data that need to be collected and the expertise needed to interpret those signals render many types of physiological signals unsuitable for common interactive intelligent systems [Delis et al. 2001]. While they can be very sensitive to cognitive activity, the above issues in combination with the degree of variability of physiological signals, due to external factors such as temperature and movement, have limited their usage for everyday environments [Delis et al. 2001]. The following sections will focus on GSR and pupillary dilation to show how physiological features are used to index cognitive load.

1.4.1 GSR for Cognitive Load

GSR is a measure of the conductivity of human skin, and can provide an indication of changes in the human sympathetic nervous system [Nourbakhsh et al. 2012]. Among different physiological signals, GSR (also referred to as electro-dermal activity (EDA)), is comparatively low-cost and easily-captured. The method involves measuring the electrical conductance of the skin through one or two sensor(s) usually attached to some part of the hand or foot. Skin conductivity varies with changes in skin moisture level (sweating) and can reveal changes in the sympathetic nervous system. Figure 1-8 shows an example of the GSR signal during a cognitive task.

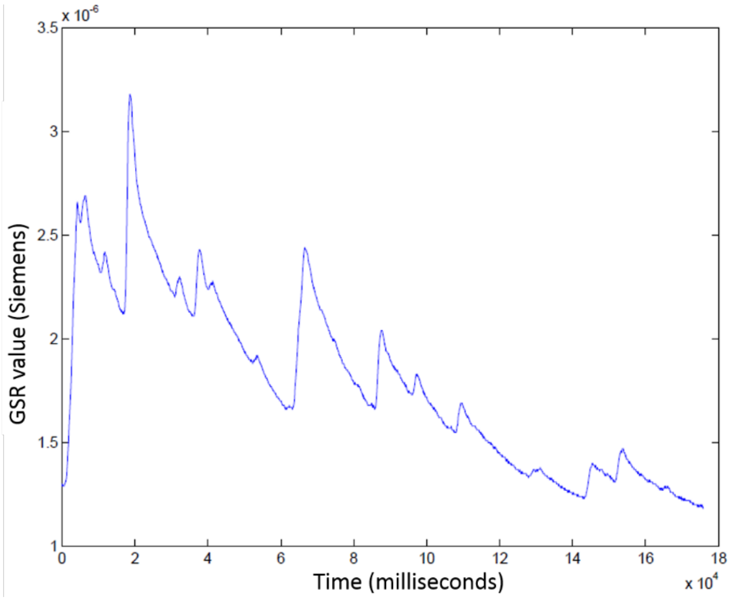


Figure 1-8. An example of the GSR signal during a cognitive task [Nourbakhsh et al. 2012].

GSR has recently been investigated in relation to mental status. Shi et al. [Shi et al. 2007] assessed GSR in stress and cognitive load situations and found correlations between GSR signals and cognitive load. It was found that GSR features such as mean GSR and accumulated GSR significantly increased when cognitive load level increased. Mean GSR is frequently used in recent work for indexing cognitive load. Wilson [Wilson 2002] analysed several physiological measures during different stages of flight in a simulator and found an increase in GSR response during take-off and landing which were the procedures expected to place the most cognitive demands on pilots. Zhou et al. [Zhou et al. 2015] used GSR to index cognitive load levels in a dynamical workload adaptation feedback loop. It showed that via manipulation of the difficulty levels, the arithmetic tasks can be adapted to match the cognitive load as indexed via GSR signals, which can improve the overall performance. Healey and Picard [Healey and Picard 2000] extracted GSR peak features such as magnitude and duration of a local peak of the GSR signal to assess car driver stress. Table 1-2 shows examples of mostly used GSR features for indexing cognitive load.

Table 1-2. Walk-through example of GSR features for indexing cognitive load.

GSR Features	**Notes**
Accumulative GSR	Accumulation of GSR values over task time

Mean of GSR	Summation of GSR values over task time divided by task time

1.4.2 Eye-based Measures for Cognitive Load

Pupillary response and eye blink have been shown to correlate with both visual and aural cognitive tasks. Eye-based measures for cognitive load have a number of advantages over the other methods. For example, eye based data collection is less intrusive than other physiological signal data collection. Pupillary response is one type of eye based workload measures and it has attracted increasing attention with the appearance of video-based eye trackers in recent years [Wilson and Schlegel 2004]. The literature on the correlation between pupillary response and cognitive workload has been abundant over the last two decades. To study task-evoked pupillary responses, Beatty [Beatty 1982] conducted experiments consisting of various tasks such as language processing, reasoning, and perception. Pupil dilation was found to be a reliable indicator of processing load during the tasks. In more recent studies in [Klingner et al. 2008; Kun et al. 2011; Chen et al. 2011] , pupillary responses were measured through video-based eye tracking and the studies found some significant eye features for the cognitive load classifications (see Table 1-3). Figure 1-9 shows an example of the pupil diameter signal with the task time setting during a cognitive task.

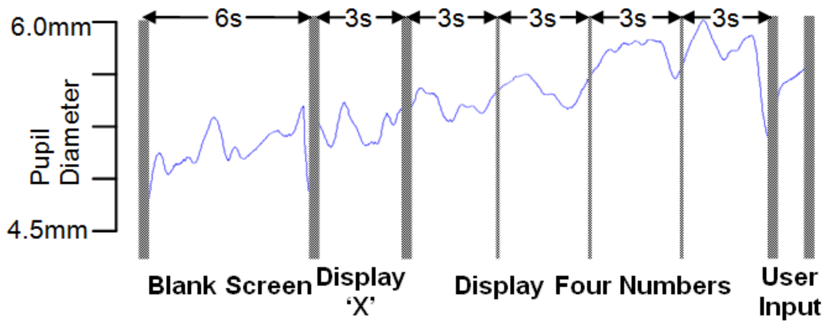


Figure 1-9. An example of the pupil diameter signal with the task time setting during a cognitive task [Wang et al. 2013].

Although there is little doubt that pupillary response is a useful measure of variations in cognitive load, pupillary response is also found to be sensitive to confounding (or noisy) factors unrelated to the cognitive task, such as changes of illumination and emotional states [Marshall 2002; Stanners et al. 1979; Xu et al. 2011] . Kramer [Kramer 1991] reported that larger changes in the pupil diameter occur in response to illumination changes than during information processing. Pomplun and Sunkara [Pomplun and Sunkara 2003] investigated the effects and the interaction of cognitive load and luminance changes in a gaze-controlled human-computer interaction task. There were three difficulty levels and two levels of background brightness (black and white) in the task. The experimental result showed that pupil size was significantly influenced by the factor of task difficulty and the factor of background brightness. However there was no interaction between the two factors.

Xu et al. [Xu et al. 2011] proposed a fine-grained feature extraction method for cognitive load measurement to deal with luminance changes in pupillary response data. In their work, the whole task period was divided into several intervals corresponding to different stages of the cognitive task as shown in Figure 1-9. As the pupil size is very sensitive to the external illumination, during each interval the illumination change was considered and normalized pupil diameter was employed to overcome the impact of external illumination for cognitive load measurement.

Chen and Epps [Chen and Epps 2013] investigated three types of eye-based approaches for CLM: pupillary response, blinking, and eye movement (fixations and saccades). Eye-based features were investigated in the presence of emotion interference which was induced by showing International Affective Picture System (IAPS) [Lang et al. 1997] image in the task background. An experiment with arithmetic-based tasks demonstrated that arousal effects were dominated by cognitive load during task execution. A set of features such as zero crossing count of pupil size, features from cumulative blink/fixation/saccade numbers, eye features from task stimuli onset to the first saccade were all proposed for cognitive load level prediction. The performance of cognitive load level prediction was found to be close to that of a reaction time measure, showing the feasibility of eye activity features for near real-time CLM.

Wang et al. [Wang et al. 2013] analysed pupillary features for cognitive load in arithmetic tasks, where each subject was seated in an office environment and performed arithmetic tasks under different luminance conditions on the monitor screen. The research found that under the same difficulty level the pupil diameter decreased with the increase of the brightness level. Furthermore, under the same brightness level, pupil diameter had a tendency to increase with task difficulty level, although the effect was not significant in this study. The study also revealed that luminance conditions had a greater effect size than cognitive demands in pupil diameter changes.

Investigations also found the relations between eye blink and cognitive load. For example, Veltman and Gaillard [Veltman and Gaillard 1998] found that the eye blink interval became shorter when the difficulty of the task increased. Ledger [Ledger 2013] found that blink rate decreased with the increase of cognitive load.

Table 1-3 shows some examples of eye features for indexing cognitive load [Veltman and Gaillard 1998; Xu et al. 2011; Wang et al. 2013; Chen and Epps 2013].

Table 1-3. Walk-through examples of eye features for indexing cognitive load.

Eye Features	Notes
Pupil diameter mean	Pupil diameter mean over the task time
Pupil diameter variance	Pupil diameter variance over the task time
Blink interval	Time between two consecutive blinks
Blink number	Number of blinks over the task time
Blink rate	Blink number divided by the task time

1.5 Multimodal Signals and Data Fusion

As discussed in [Oviatt and Cohen 2015], multimodal interfaces have been shown to enhance human-computer interaction by allowing users to communicate more naturally and interact with complex information with more freedom of expression than traditional computer interfaces. Therefore, it is natural to investigate how cognitive load can be examined from multiple modalities in order to get robust cognitive load measurement. This section firstly introduces theories related to multimodal cognitive load and an abstract model for multimodal cognitive load measurement is presented for robust cognitive load measurement.

1.5.1 Multimodality and Cognitive Load

During the past decade, Oviatt and colleagues began investigating the relation between multimodal interaction and cognitive load in users [Oviatt et al. 2004; Oviatt 2006; Ruiz et al. 2010b]. The close relationship between multimodality and cognition exists on a number of levels [Ruiz et al. 2010b]. First, multimodal processing based on psychological and cognitive theories, as well as neurological evidence, provide a biological and theoretical perspective to explain the advantages of multimodal interaction [James et al. 2017; Oviatt 2017]. Previous work has shown that interacting multimodally can reduce users' cognitive load, and that users actively self-manage their own cognitive load by shifting between unimodal and multimodal interaction patterns [Oviatt et al. 2004; Oviatt 2006]. Empirical evidence suggests that when tasks increasingly become more difficult users prefer to interacting multimodally rather than unimodally across a variety of different application domains [Oviatt et al. 2004]. In an educational context, interacting multimodally helps students manage their limited working memory resources as problems become harder [Oviatt et al. 2006]. Cognitive load theory [Chandler and Sweller 1991], which is based on Working Memory theory [Baddeley 1992], has provided a context for understanding why multimodal-multimedia interaction leads to performance improvements [Oviatt 2017].

Furthermore, human beings are physiologically designed to acquire and produce information through a number of different modalities: the human communication channel is made up of sensory organs, the central nervous system, various parts of the brain and effectors (muscles or glands) [Kandel et al. 2000]. Sensory inputs from specific modalities each have their own individual pathway into a primary sensory cortex and can be processed in parallel [Kandel et al. 2000]. In addition, there are specific multimodal integration (input) and diffusion (output) association areas in the brain that are highly interconnected [Kandel et al. 2000]. Adding credence to the raft of empirical evidence, neuroimaging technology such as positron emission topography and functional magnetic resonance imaging have been used to identify separate locations for the verbal/auditory, imagery/spatial and executive functions of working memory [Smith et al. 1996; Awh et al. 1996]. Multimodal perception and cognition structures in the human brain appear to have been specifically designed to collate and produce multimodal information.

Multimodal feature combinations can provide a valuable index of cognitive load. For example, symptomatic cues of cognitive load derived from user behavior such as acoustic speech signals (e.g. acoustic), transcribed text (e.g. pauses, pronouns, valence, complexity), digital pen trajectories of handwriting (e.g. velocity, length, pressure, orientation), eye activities, and skin conductance, can be supported by well-established theoretical frameworks, including Sweller's Cognitive Load Theory [Chandler and Sweller 1991], and Baddeley's model of working memory [Baddeley 1992]. Empirical evidence suggests that users prefer to interact multimodally rather than unimodally with the increase of cognitive load [Oviatt et al. 2004], which provides an opportunity to more robustly detect shifts in users' cognitive load unobtrusively during tasks. Temporal dynamics that exist between interaction modalities (e.g., speech and pen) have also been shown to change under increased load conditions [Oviatt et al. 2004]. All these suggest that it is feasible to use features extracted from multiple modal inputs for *multimodal fusion* as indices of cognitive load.

1.5.2 Walk-Through of a System for Multimodal Cognitive Load Measurement

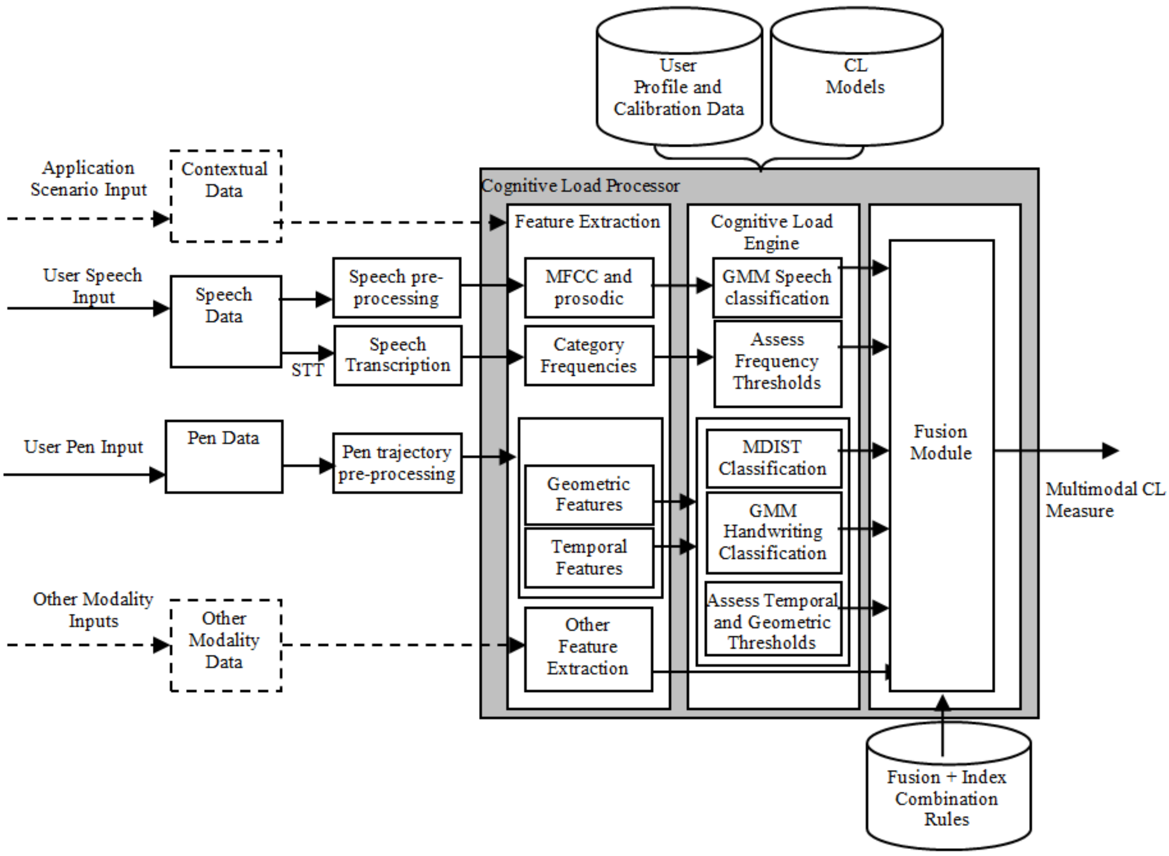


Figure 1-10. High level functional model of a multiple modality CLM system [Chen et al. 2012].

Figure 1-10 depicts a high level functional model of a proposed Multimodal Cognitive Load Measurement (MCLM) system [Chen et al. 2012]. The abstract system model embodies four high level processes: pre-processing and data cleaning, feature extraction, load assessment and index fusion. In this system, multimodal behavioral indices of cognitive load are collected during basic multimodal input interpretation, so the same information also can be routed for interpretation of the user's cognitive load level. The raw modality input sources are first and foremost intended for purposes other than cognitive load measurement, specifically to do with the domain application. For example, the data may be used for semantic interpretation or rendering (e.g. in the case of command and control speech or interactive pen gestures). The data may therefore need to be duplicated and diverted – with the original stream sent to the recognizers, and a secondary stream sent to the Cognitive Load Measurement engine. In Figure 1-10 , speech input data is first captured through a close-talk microphone. This generates two kinds of data, speech signal data (e.g. a wav file) and text (through a speech to text engine). Likewise, pen input data is collected as trajectory tuples, including pressure, pen orientation and other information transmitted directly from the device drivers, alongside system timestamps.

Data pre-processing and data cleaning refers to any reformatting, restructuring of the input data, or removal of unnecessary information, for example, any outliers or segments that are too short for geometric (for writing trajectory features) and temporal analysis; words not recognized in the text, as well as words that are not used in the analysis. Input streams from other modalities will follow the same processes. Similarly, a number of other non-behavioral indices will also undergo pre-processing as needed; these include indices that may also be used in the process, such as galvanic skin response, or other body-based data, such as posture, movement or temperature. Environmental and other external context information may also be provided to the MCLM system for enhanced performance at this point.

The second stage involves streaming the individual modal input into their respective feature extraction components. The same data may be used for multiple feature extraction components, while other extraction components may not be activated, depending on domain-specific contextual information gathered from the active applications and workflow diagrams established a-priori. This will allow the feature extraction engine to choose the most appropriate modules to activate for each incoming input stream. For example, if the incoming speech is sourced from a phone call or radio conversation, the feature extraction component will activate both MFCC and prosodic feature extraction as well as the linguistic category extraction components, since both can provide meaningful measures on this kind of data. On the other hand, if the incoming speech is sourced from command and control input, only MFCC and prosodic feature extraction will be activated, as the linguistic categories cannot provide any meaningful cognitive load measurement information on short, closed vocabulary, single word speech. Types of speech features can vary from intensity, pitch, and formants to other acoustic prosodic, or linguistic features such as grammatical features, pause, language complexity as well as word categories and valence. Examples of writing features are writing velocity and acceleration, stroke length, pressure, writing orientation, and frequency.

The third stage involves the decision-making aspect of the process, where thresholds are invoked and the appropriate models for each modality are selected from the database from which to carry out the classification. For example, for the speech signal based cognitive load measurement, different models are required for single word cognitive load classification versus continuous speech classification. Likewise, different MDIST models exist for each shape, and also for each user. Any calibration data that is needed for classification or for comparison purposes is also accessed at this point.

The final stage involves the fusion of indices resolved from the previous stage. The assessment results obtained from each modality can also convey confidence information to support the fusion process. The fusion engine accesses information regarding the modality load assessment combination rules in each specific context, e.g. whether the time-windows for the collected inputs are compatible; which indices are complementary with which others; and the appropriate weightings for each index, given the scenario and the user situation. Figure 1-10 shows how mid- and late- fusion may be achieved from a set of cognitive load assessments from each of the sub-features. Mid-fusion refers to the fusion of features extracted from multimodalities before cognitive load classification, while late-fusion is the fusion of classification scores from single modality decisions (see Figure 1-11).

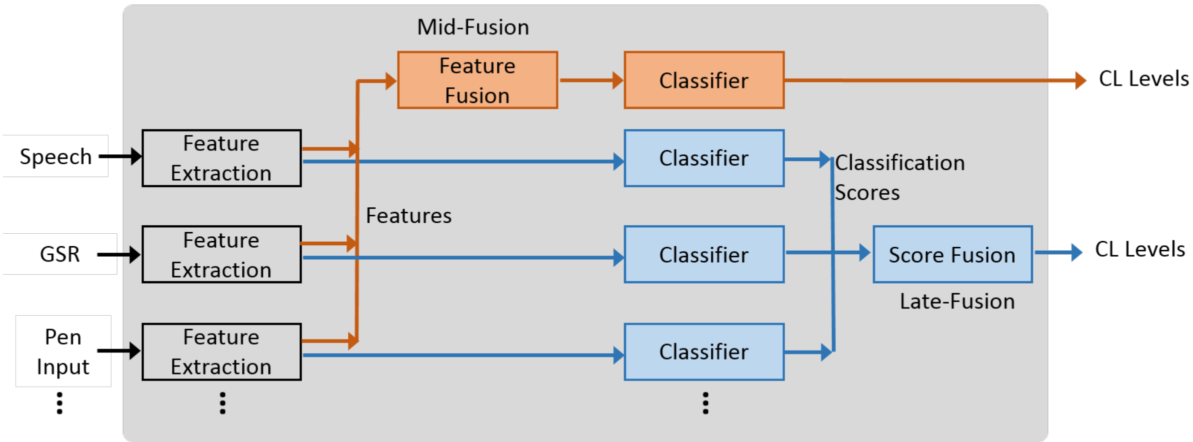


Figure 1-11. Fusion strategies of mid-fusion and late-fusion.

1.5.3 Case Study: Cognitive Load during Cognitive Skill Training

In order to illustrate how a multimodal cognitive load measurement system works, a lab-based study is presented in which cognitive load and complexity were manipulated, and multiple behavioral modalities were recorded. The objective is to assess how well individual and combined modalities can reflect levels of cognitive load, and provide a concrete application for the multimodal cognitive load measurement model. A cognitive skill training experiment was adopted for the examination, conducted by elite athletes from the Australian Institute of Sport (AIS). This experiment was adopted due to the following considerations. First, the athletes have enough knowledge regarding sports which could minimize the pre-experiment training. Second, cognitive skill is an important part of their everyday training, which meanwhile provides us the unique opportunity to examine cognitive load with realistic data. Finally, typical interaction modalities such as pen input, can be conveniently embedded into their sport skill training scheme.

Specifically in the experiment, participants were required to complete cognitive skills training using a targeted sports-specific software application called AISReact [Mackintosh 2010]. While aiming at ever faster situation analysis and decision making through the construction better mental schemas, it is desirable to precisely determine onsets of very high cognitive load in order to adapt the training rate to each individual athlete. In this experiment, we modified the software to accept pen based interaction, and added the modalities of speech and eye-activity. In addition, performance (accuracy) measures, GSR and subjective ratings were also collected to establish a ground truth for cognitive load and task difficulty [Chen et al. 2012]. The setup is shown in Figure 1-12.

Figure 1-12. Physical set-up of a user completing a task using a digital pen and with GSR attached.

(a)

(b)

Figure 1-13. Video clips of basketball plays example in the study: (a) Last frame of video clip before freeze, (b) Blank court image with player markings.

Twelve male recreational basketball players, aged 19-36, each with more than 2 years' experience (average of 9.4 years) volunteered to complete the study. The task consisted of a 10-second video basketball clip played on a tablet monitor, which was then frozen and replaced with a blank court schematic. The clips involved 10 players and the participants had to remember the locations and roles of some players in three task difficulty levels (remember 3 players for Low level, 6 for Medium, and all 10 for High). Each level consisted of 6 distinct clips. The clips were filmed from above and cover half the court, with all plays moving from the bottom of the screen towards the top, where the basketball hoop was located, as seen in Figure 1-13.

Participants used specific pen marks to identify the remembered player positions on the tablet monitor: attackers were denoted by crosses, defenders by circles and the ball carrier by a circle with a dot in the middle, as illustrated in Figure 1-13. Participants were also instructed to think aloud through their answers, and these utterances were captured using a close-talk microphone. The speech, GSR and pen input signals were collected during the experiment.

For cognitive load classification, the speech data yielded a classification accuracy of 61.8% for identifying the three levels of cognitive load [Ruiz et al. 2010a] using the rate of pitch peaks as features. Specifically, the rate of pitch peaks increases as higher cognitive load is experienced by the speaker. In comparison, Table 1-4 summarizes the pen signal features and their individual accuracies for classifying load levels. The results range from 31% to 41% for this 3-level classification, in some cases not outperforming chance classification of 33.3%. Finally, GSR signal was simply analyzed using mean GSR over the task period, yielding a classification accuracy of 64.4% over 3 load levels, across all 9 subjects, using a leave-one out evaluation scheme..

Table 1-4. Pen-input trajectory features.

Geometric feature	Description	Accuracy on test samples
Duration	Stroke duration, in milliseconds	**32.6%**
Length	Cumulative distance between sampled points along the trajectory	**40.7%**
Mean velocity	Mean velocity of the stroke trajectory, calculated point to point	**30.7%**
Mean acceleration	Mean acceleration of the stroke trajectory, calculated point to point	**37.0%**
Area	The area in pixels taken by the circle shape, enclosed by the trajectory	**36.3%**
First.Last	Distance between the first and last points of the trajectory	**33.3%**
Overlap ratio	The ratio of the overlapping distance between the first and last points of the trajectory to the total size of the shape.	**37.4%**

In this case study, the speech feature in terms of rate of pitch peaks as presented in the previous section, the pen input features and GSR are fused using the AdaBoost boosting algorithm (late-fusion). Boosting [Freund 1995; Schapire 1990] is a general ensemble learning algorithm, which creates an accurate strong classifier H by iteratively combining a number T of moderately inaccurate weak classifiers h_t. By definition, a strong classifier has high classification accuracy on the data set, while a weak classifier's accuracy is just above that of a random guess. The final strong classifier can be defined as:

$$H(x) = \begin{cases} 1 & \text{if } \sum_{t=1}^{T} \alpha_t h_t(x) \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

where α_t is a weight coefficient. In simple cases, each weak classifier is attached to a feature, so the process of combining weak classifiers in Boosting is equivalent to a feature fusion process.

Table 1-5 details the weights obtained for the fusion of speech and the pen features. The average classification accuracy when fusing all these features is 64.1% on the testing samples, for the 3 load levels across all 9 subjects. It should be noted that this represents a small improvement over the speech-only accuracy in the previous section.

Table 1-5. AdaBoost weights for speech and pen input features.

	Pitch Peaks	**Duration**	**Length**	**Velocity**	**Acceleration**	**Area**	**First-Last**	**Overlap**
Weights	0.686	0.150	0.053	0.000	0.051	0.059	0.000	0.002

Similarly, Table 1-6 details the weights obtained when fusing speech, pen features and GSR. The average classification accuracy is improved to 77.8% on the testing samples, for the 3 load levels across all 9 subjects.

Table 1-6. AdaBoost weights for speech, pen input features and GSR.

	Pitch Peaks	Duration	Length	Velocity	Acceleration	Area	First-Last	Overlap	GSR
Weights	0.478	0.176	0.055	0.05	0.041	0.053	0.011	0.000	0.181

Adding the GSR feature provides a significant improvement, supporting the benefits of feature fusion for cognitive load classifications. This case study is proposed as one implementation example of the model, however the results indicate that other behavioral features, yet to be explored, may be able to provide further multimodal cognitive load measurement accuracy.

1.6 Conclusion

Cognitive load is a multi-dimensional construct representing the load imposed on the working memory during performance of a cognitive task. It is a critical factor impacting human performance in modern complex mission-critical domains such as aviation, and even military command and control. Different approaches have been developed for cognitive load measurement ranging from subjective ratings, performance measures, and behavioral measures to physiological measures. This chapter investigated multimodal behavioral and physiological signals as indicators of cognitive load, due to their potential for real-time, non-intrusive cognitive load monitoring. In specific, features from behavioral modalities such as pen input, speech were reviewed for cognitive load measurement, and similarly physiological features such as GSR and eye-based features were discussed as indicators of cognitive load. *Multimodal fusions* were finally presented as a robust multimodal cognitive load measurement.

Based on the multimodal cognitive load measurement, it is possible to dynamically adjust task options in real-time in order to keep users in an optimal cognitive state and improve their engagement and performance. We bring forward the prospect of MCLM in the near future, however many new research and development opportunities may arise beyond them:

- Real-time cognitive load measurement with the dynamic fusion of multimodalities. The work discussed in this chapter mainly focuses on the offline processing of signals from multimodalities for cognitive load measurement, however real-time processing is the key to enable the dynamic task adjustment based on the cognitive load experience. While the dynamic fusion of multimodalities allows the use of modalities for the cognitive load measurement dynamically (e.g. unimodality, multimodalities, or just some of features from unimodality or multimodalities) based on each modality properties, task scenarios, subject characteristics, or other information. Therefore, robust multimodal cognitive load measurement is set up for the accurate evaluation of human cognitive states.
- Dynamic cognitive load adjustment in a feedback loop. This requires a setup of adaptation model of cognitive load so that task properties and other attributes such as the work environment related to the user are updated automatically for an optimal cognitive state of the user.
- Modelling cognitive load with more advanced approaches. We imagine that future work on MCLM will include furthering the cause of Bayesian models [Griffiths et al. 2008; Austerweil et al. 2015] of cognition (as well as the more recently introduced Quantum models [Bruza et al. 2015] of cognition) by providing them with a framework that employs empirical/sensory links to external tangible world. Such theoretical models of cognition hold much promise for modelling the possible mental states and the transitions between them. Data-driven MCLM can provide empirical associations of such mental states and corresponding transformations to various user interaction modality patterns.
- New technologies and devices for cognitive load measurement. Many current techniques, such as GSR, digital pen, eye tracker etc. have been applied in the examinations of this chapter. However, there is still much space to explore and improve on for cognitive load measurement, likely via employing new technologies and devices, such as virtual reality methods for experiments, big data analytics for group experiences of cognitive load, longitudinal tracking to construct cognitive profile for a single subject, and experience-based cognitive load studies.

In summary, the real-time cognitive load measurement offers new potential for dynamic support and adaptive system behavior, promising to optimize the human-machine interaction throughput, and therefore improve human engagement and performance by reducing the burden placed on limited human cognitive capabilities.

Acknowledgements

The material in this chapter is based in part on work supported by the Asian Office of Aerospace Research & Development (AOARD) under grant No. 134131, 134144, 124049, 114056.

Focus Questions

1. What constitutes the cognitive load construct, and how does it affect the performance of people?
2. How can cognitive load be measured, and what are the respective strengths and limitations for different means of cognitive load measurement?
3. What factors may impact human's cognitive load, or cause bias in the cognitive load measurement?
4. The limited capacity of working memory, as a determinant of human performance, is considered important in human learning. How can a learner's learning process, based on this theory, be tailored to improve the learning outcome?
5. What is typical behavioral methods for cognitive load measurement, and why are they considered suitable for real-time cognitive load examination?
6. Physiological methods can be used to measure cognitive load using signals such as EEG or GSR. Under what contexts can physiological methods be applied for cognitive load measurement, and how do their features differ from behavioral methods?
7. When available, multimodal signals are expected to improve the accuracy of cognitive load measurement. How is this achieved in realistic scenarios, and what is the commonly applied strategy of data fusion?
8. What is the main difference between the multimodal cognitive load measurement system and a single modality cognitive load measurement system, and why this difference occurs?
9. Using the current knowledge introduced in this chapter, what factors should be considered in designing a cognitive load measurement system?
10. Design a simple procedure that will: induce different levels of cognitive load, collect user's hand motion and GSR data, and measure the cognitive load of the user.

References

Ritu Agarwal and Elena Karahanna. 2000. Time Flies when You're Having Fun: Cognitive Absorption and Beliefs About Information Technology Usage1. *MIS Quarterly* 24, 4 (December 2000), 665–694. DOI:https://doi.org/10.2307/3250951

Gilles Allali, Frederic Assal, Reto W. Kressig, Veronique Dubost, Francois R. Herrmann, and Olivier Beauchet. 2008. Impact of impaired executive function on gait stability. *Dementia and Geriatric Cognitive Disorders* 26, 4 (2008), 364–369.

Syed Z. Arshad, Yang Wang, and Fang Chen. 2013. Analysing Mouse Activity for Cognitive Load Detection. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application,* *Innovation, Collaboration*. OzCHI '13. New York, NY, USA: ACM, 115–118. DOI:https://doi.org/10.1145/2541016.2541083

Joseph L. Austerweil, Samuel J. Gershman, and Thomas L. Griffiths. 2015. Structure and Flexibility in Bayesian Models of Cognition. In James T. Townsend & Jerome R. Busemeyer, eds. *The Oxford Handbook of Computational and Mathematical Psychology*. Oxford University Press.

Edward Awh, John Jonides, Edward E. Smith, Eric H. Schumacher, Robert A. Koeppe, and Stewart Katz. 1996. Dissociation of Storage and Rehearsal in Verbal Working Memory: Evidence From Positron Emission Tomography. *Psychological Science* 7, 1 (January 1996), 25–31. DOI:https://doi.org/10.1111/j.1467-9280.1996.tb00662.x

Paul Ayres and Fred Paas. 2009. Interdisciplinary Perspectives Inspiring a New Generation of Cognitive Load Research. *Educ Psychol Rev* 21, 1 (March 2009), 1–9. DOI:https://doi.org/10.1007/s10648-008-9090-7

Jonathan Back and Charles Oppenheim. 2001. A model of cognitive load for IR: implications for user relevance feedback interaction. *Information Research* 6, 2 (January 2001).

R.W. Backs and L.C. Walrath. 1992. Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Applied Ergonomics* 23 (1992), 243–254.

A.D. Baddeley. 1992. Working Memory. *Science* 255 (1992), 556–559.

Alan Baddeley. 2012. Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology* 63, 1 (2012), 1–29. DOI:https://doi.org/10.1146/annurev-psych-120710-100422

J. Beatty. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol Bull* 91, 2 (March 1982), 276–292.

A. Berthold and A. Jameson. 1999. Interpreting Symptoms of Cognitive Load in Speech Input. In *Seventh International Conference on User Modeling (UM99)*. Springer Wien New York.

R. Brunken, J.L. Plass, and D. Leutner. 2003. Direct measurement of cognitive load in multimedia learning. *Educational Psychologist* 38, 1 (2003), 53–61.

Peter D. Bruza, Zheng Wang, and Jerome R. Busemeyer. 2015. Quantum cognition: a new theoretical approach to psychology. *Trends in Cognitive Sciences* 19, 7 (January 2015), 383–393. DOI:https://doi.org/10.1016/j.tics.2015.05.001

Janis A. Cannon-Bowers, Eduardo Salas, and Sharolyn Converse. 1993. Shared mental models in expert team decision making. In J. Castellan, ed. *Individual and Group Decision-Making: Current Issues*. Hillsdale, NJ: Lawrence Erlbaum Associates, 221–246.

Paul Chandler and John Sweller. 1991. Cognitive Load Theory and the Format of Instruction. *Cognition and instruction* 8, 4 (1991), 293–332.

Fang Chen et al. 2012. Multimodal Behavior and Interaction As Indicators of Cognitive Load. *ACM Transactions on Interactive Intelligent Systems* 2, 4 (December 2012), 22:1–22:36.

Fang Chen et al. 2016. *Robust Multimodal Cognitive Load Measurement*, Cham: Springer International Publishing.

Siyuan Chen and Julien Epps. 2013. Automatic classification of eye activity for cognitive load measurement with emotion interference. *Comput Methods Programs Biomed* 110, 2 (May 2013), 111–124. DOI:https://doi.org/10.1016/j.cmpb.2012.10.021

Siyuan Chen, Julien Epps, Natalie Ruiz, and Fang Chen. 2011. Eye Activity As a Measure of Human Mental Effort in HCI. In *Proceedings of the 16th International Conference on Intelligent User Interfaces*. IUI '11. New York, NY, USA: ACM, 315–318. DOI:https://doi.org/10.1145/1943403.1943454

Gabriele Cierniak, Katharina Scheiter, and Peter Gerjets. 2009. Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior* 25, 2 (March 2009), 315–324. DOI:https://doi.org/10.1016/j.chb.2008.12.020

N. Cowan. 2001. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 24, 1 (2001), 87–114.

David E. Crundall and Geoffrey Underwood. 1998. Effects of experience and processing demands on visual information acquisition in drivers. *Ergonomics* 41, 4 (April 1998), 448–458. DOI:https://doi.org/10.1080/001401398186937

Nicolas Debue and Cécile van de Leemput. 2014. What does germane load mean? An empirical contribution to the cognitive load theory. *Front. Psychol* 5 (2014), 1099. DOI:https://doi.org/10.3389/fpsyg.2014.01099

D.C. Delis, J.H. Kramer, and E. Kaplan. 2001. *The Delis-Kaplan Executive Function System*, The Psychological Corporation.

M. Dindar, I. Kabakçı Yurdakul, and F. İnan Dönmez. 2014. Measuring cognitive load in test items: static graphics versus animated graphics. *Journal of Computer Assisted Learning* 31, 2 (November 2014), 148–161.

Johan Engström, Emma Johansson, and Joakim Östlund. 2005. Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Traffic Psychology and Behaviour* 8, 2 (March 2005), 97–120. DOI:https://doi.org/10.1016/j.trf.2005.04.012

E. Ferreira et al. 2014. Assessing real-time cognitive load based on psycho-physiological measures for younger and older adults. In *2014 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*. 39–48. DOI:https://doi.org/10.1109/CCMB.2014.7020692

Yoav Freund. 1995. Boosting a Weak Learning Algorithm by Majority. *Information and Computation* 121, 2 (September 1995), 256–285. DOI:https://doi.org/10.1006/inco.1995.1136

V.J. Gawron. 2000. *Human Performance Measures Handbook*, New Jersey: Lawrence Erlbaum Associates.

Tamara van Gog, Fred Paas, Nadine Marcus, Paul Ayres, and John Sweller. 2008. The Mirror Neuron System and Observational Learning: Implications for the Effectiveness of Dynamic Visualizations. *Educ Psychol Rev* 21, 1 (November 2008), 21–30. DOI:https://doi.org/10.1007/s10648-008-9094-3

D. Gopher and R. Braune. 1984. On the psychophysics of workload: Why bother with subjective measures? *Human Factors* 26 (1984), 519–532.

Thomas Griffiths, Charles Kemp, and Joshua Tenenbaum. 2008. Bayesian models of cognition. In Ron Sun, ed. *The Cambridge Handbook of Computational Psychology*. Cambridge University Press.

S. Hart and L. Staveland. 1988. Development of the NASA-TLX (Task Load Index): Results of Empirical and Theoretical research. In P. A. Hancock & N. Meshkati, eds. *Human Mental Workload*. Elsevier.

J. Healey and R. Picard. 2000. SmartCar: detecting driver stress. In *Proceedings of 15th International Conference on Pattern Recognition 2000*. 218–221.

Karen L. Hessler and Ann M. Henderson. 2013. Interactive Learning Research: Application of Cognitive Load Theory to Nursing Education. *International Journal of Nursing Education Scholarship* 10, 1 (January 2013). DOI:https://doi.org/10.1515/ijnes-2012-0029

Paul Jen-Hwa Hu, Pai-Chun Ma, and Patrick Y.K. Chau. 1999. Evaluation of user interface designs for information retrieval systems: a computer-based experiment. *Decision Support Systems* 27, 1–2 (November 1999), 125–143. DOI:https://doi.org/10.1016/S0167-9236(99)00040-8

C.S. Ikehara and M.E. Crosby. 2005. Assessing Cognitive Load with Physiological Sensors. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*. 295a–295a. DOI:https://doi.org/10.1109/HICSS.2005.103

S.T. Iqbal, X.S. Zheng, and B.P. Bailey. 2004. Task-Evoked Pupillary Response to Mental Workload in Human-Computer Interaction. In *Proceedings of the Internatinal Conference on Computer-Human Interaction (CHI 2004)*. Vienna, Austria.

S.C. Jacobs et al. 1994. Use of skin conductance changes during mental stress testing as an index of autonomic arousal in cardiovascular research. *The American Heart Journal* 128, 1 (1994), 6.

K. James, S. Vinci-Booher, and F. Munoz-Rubke. 2017. The impact of multimodalmultisensory learning on human performance and brain activation patterns. In S. Oviatt, B. Schuller, P. Cohen, D. Sonntag, G. Potamianos, & A. Krueger, eds. *The Handbook of Multimodal-Multisensor Interfaces, Volume 1: Foundations, User Modeling, and Common Modality Combinations*. San Rafael, CA: Morgan Claypool Publishers.

Eric Kandel, James Schwartz, and Thomas Jessell. 2000. *Principles of Neural Science* 4 edition., New York: McGraw-Hill Medical.

D. Kennedy and A. Scholey. 2000. Glucose administration, heart rate and cognitive performance: effects of increasing mental effort. *Psychopharmacology* 149, 1 (2000), 63–71.

H. Keraenen et al. 2004. Prosodic Features of Speech Produced By Military Pilots During Demanding Tasks. In Finland Oulu, ed. *Proceedings of the Fonetiikan Paeivaet 2004*.

B. Kerr. 1973. Processing demands during mental operations. *Memory and Cognition* 1 (1973), 401–412.

M. Asif Khawaja, Fang Chen, and Nadine Marcus. 2014. Measuring Cognitive Load Using Linguistic Features: Implications for Usability Evaluation and Adaptive Interaction Design. *International Journal of Human-Computer Interaction* 30, 5 (May 2014), 343–368. DOI:https://doi.org/10.1080/10447318.2013.860579

Jeff Klingner, Rakshit Kumar, and Pat Hanrahan. 2008. Measuring the Task-evoked Pupillary Response with a Remote Eye Tracker. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*. ETRA '08. New York, NY, USA: ACM, 69–72. DOI:https://doi.org/10.1145/1344471.1344489

A.F. Kramer. 1991. Physiological metrics of mental workload: a review of recent progress. In D. L. Damos, ed. *Multiple-task performance*. London: Taylor and Francis, 279–328.

Seffetullah Kuldas, Lata Satyen, and Hairul Ismail. 2014. Greater Cognitive Effort for Better Learning: Tailoring an Instructional Design for Learners with Different Levels of Knowledge and Motivation. *Psychologica Belgica* 54, 4 (August 2014), 350–373. DOI:https://doi.org/10.5334/pb.aw

Andrew L. Kun, Zeljko Medenica, Oskar Palinko, and Peter A. Heeman. 2011. Utilizing Pupil Diameter to Estimate Cognitive Load Changes During Human Dialogue: A Preliminary Study. In *AutomotiveUI 2011 Adjunct Proceedings*. Salzburg, Austria.

P.J. Lang, M.M. Bradley, and B.N. Cuthbert. 1997. *International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual*, Florida, USA: University of Florida.

Hayley Ledger. 2013. The effect cognitive load has on eye blinking. *The Plymouth Student Scientist* 6, 1 (2013), 206–223.

Jimmie Leppink, Fred Paas, Cees P.M. Van der Vleuten, Tamara Van Gog, and Jeroen J.G. Van Merriënboer. 2013. Development of an instrument for measuring different types of cognitive load. *Behav Res Methods* 45, 4 (December 2013), 1058–1072. DOI:https://doi.org/10.3758/s13428-013-0334-1

O.V. Lipp and D.L. Neumann. 2004. Attentional blink reflex modulation in a continuous performance task is modality specific. *Psychophysiology* 41, 3 (2004), 417–425.

C. Mackintosh. 2010. *AIS React Software v.6.6*, Australian Institute of Sport.

N. Marcus, M. Cooper, and J. Sweller. 1996. Understand Instructions. *Educational Psychology* 88, 1 (1996), 49–63.

S.P. Marshall. 2002. The Index of Cognitive Activity: measuring cognitive workload. In *Proceedings of the 2002 IEEE 7th Conference on Human Factors and Power Plants, 2002*. 7-5-7–9. DOI:https://doi.org/10.1109/HFPP.2002.1042860

S.P. Marshall, C.W. Pleydell-pearce, and B.T. Dickson. 2003. Integrating psychological measures of cognitive workload and eye movements to detect strategy shifts. In *Proceedings of 36th Hawaii International Conference on System Sciences (HICSS 2003)*. IEEE Computer Society.

E. Martin and R. Bajcsy. 2011. Leveraging Wireless Sensors and Smart Phones to Study Gait Variability. In Azizah Abd Manaf, Shamsul Sahibuddin, Rabiah Ahmad, Salwani Mohd Daud, & Eyas El-Qawasmeh, eds. *Informatics Engineering and Information Science*. Communications in Computer and Information Science. Springer Berlin Heidelberg, 95–111.

Jorge Moll, Roland Zahn, Ricardo de Oliveira-Souza, Frank Krueger, and Jordan Grafman. 2005. The neural basis of human moral cognition. *Nat Rev Neurosci* 6, 10 (October 2005), 799–809. DOI:https://doi.org/10.1038/nrn1768

Briana B. Morrison, Brian Dorn, and Mark Guzdial. 2014. Measuring Cognitive Load in Introductory CS: Adaptation of an Instrument. In *Proceedings of the Tenth Annual Conference on International Computing Education Research*. ICER '14. ACM, 131–138. DOI:https://doi.org/10.1145/2632320.2632348

S.Y. Mousavi, R. Low, and J. Sweller. 1995. Measurement and analysis methods of heart rate and respiration for use in applied environments. *Journal of Educational Psychology* 87, 2 (1995), 319–334.

C. Mueller, B. Grossmann-hutter, A. Jameson, R. Rummer, and F. Wittig. 2001. Recognising time pressure and cognitive load on the basis of speech: An experimental study. In *Proceedings of the Eighth International Conference on User Modeling (UM 2001)*. Berlin.

Kyoungsik Na. 2012. *Exploring The Effect Of Cognitive Load On The Propensity For Query Reformulation Behavior*. PhD Thesis. The Florida State University.

P. Nickel and F. Nachreiner. 2000. Psychometric Properties of the 0.1Hz Component of HRV as an Indicator of Mental Strain. In *Proceedings of the IEA 2000/HFES 2000: the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society*. San Diego, California.

Nargess Nourbakhsh, Yang Wang, Fang Chen, and Rafael A. Calvo. 2012. Using Galvanic Skin Response for Cognitive Load Measurement in Arithmetic and Reading Tasks. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*. OzCHI '12. 420–423. DOI:https://doi.org/10.1145/2414536.2414602

R. O'Donnell and F. Eggemeier. 1986. Workload assessment methodology. In K. Boff, L. Kaufman, & J. Thomas, eds. *Handbook of Perception and Human Performance. Cognitive Processes and Performance*. Wiley, 42–1–42–49.

S. Oviatt. 2006. Human-Centered Design Meets Cognitive Load Theory: Designing Interfaces that Help People Think. In *Proceedings of the 14th annual ACM international conference on Multimedia (MULTIMEDIA 2006)*.

S. Oviatt, R. Coulston, and R. Lunsford. 2004. When do we interact multimodally?: Cognitive load and multimodal communication patterns. In *Proceedings of the 6th international conference on Multimodal interfaces (ICMI 2004)*. New York, USA.

Sharon Oviatt. 2013. *The Design of Future Educational Interfaces*, New York: Routledge.

Sharon Oviatt. 2017. Theoretical Foundations of Multimodal Interfaces and Systems. In S. Oviatt, B. Schuller, P. Cohen, D. Sonntag, G. Potamianos, & A. Krueger, eds. *The Handbook of Multimodal-Multisensor Interfaces, Volume 1: Foundations, User Modeling, and Common Modality Combinations*. San Rafael, CA: Morgan Claypool Publishers.

Sharon Oviatt, Alex Arthur, and Julia Cohen. 2006. Quiet Interfaces That Help Students Think. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*. UIST '06. New York, NY, USA: ACM, 191–200. DOI:https://doi.org/10.1145/1166253.1166284

Sharon Oviatt and Philip R. Cohen. 2015. The Paradigm Shift to Multimodality in Contemporary Computer Interfaces. In Jack Carroll, ed. *Synthesis Lectures on Human-Centered Informatics*. San Rafael, CA: Morgan Claypool Publishers.

F. Paas, P. Ayers, and M. Pachman. 2008. Assessment of cognitive load in multimedia learning: Theory, methods and applications. In D. H. Robinson & G. Schraw, eds. *Recent Innovations in Educational Technology that Facilitate Student Learning*. Information Age Publishing Inc, 11–36.

Fred Paas and Jeroen Merriënboer. 1994. Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review* 6, 4 (December 1994), 351–371. DOI:https://doi.org/10.1007/BF02213420

Fred Paas, Juhani E. Tuovinen, Huib Tabbers, and Pascal W.M. Van Gerven. 2003. Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist* 38, 1 (2003), 63–71. DOI:https://doi.org/10.1207/S15326985EP3801_8

Ngoc Le Phu. 2011. *The use of spectral information in the development of novel techiques for speech-based cognitive load classification*. University of New South Wales.

Marc Pomplun and Sindhura Sunkara. 2003. Pupil dilation as an indicator of cognitive workload in human-computer interaction. In *Proceedings of the International Conference on HCI 2003*.

N. Ruiz, G. Liu, B. Yin, D. Farrow, and F. Chen. 2010a. Teaching Athletes Cognitive Skills: Detecting Load in Speech Input. In *Proceedings of the 24th BCS Conference on Human Computer Interaction (HCI 2010)*. Dundee, Scotland.

Natalie Ruiz, Fang Chen, and Sharon Oviatt. 2010b. Multimodal Input. In Jean-Philippe Thiran, Ferran Marqués, & Hervé Bourlard, eds. *Multimodal Signal Processing*. Elsevier.

R.E. Schapire. 1990. The strength of weak learnability. *Machine Learning* 5, 2 (1990), 197–227.

Kilian G. Seeber. 2013. Cognitive load in simultaneous interpreting: Measures and methods. *Target* 25, 1 (January 2013), 18–32. DOI:https://doi.org/10.1075/target.25.1.03see

J. Bryan Sexton and Robert L. Helmreich. 2000. Analyzing Cockpit Communication: The Links between Language, Performance, Error, and Workload. *The Journal of the Human Performance in Extreme Environments* 5, 1 (2000), 63–68.

Yu Shi, Eric Choi, Ronnie Taib, and Fang Chen. 2010. Designing Cognition-Adaptive Human–Computer Interface for Mission-Critical Systems. In George Angelos Papadopoulos, Wita Wojtkowski, Gregory Wojtkowski, Stanislaw Wrycza, & Jože Zupancic, eds. *Information Systems Development*. Springer US, 111–119.

Yu Shi, Natalie Ruiz, Ronnie Taib, Eric Choi, and Fang Chen. 2007. Galvanic Skin Response (GSR) As an Index of Cognitive Load. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '07. New York, NY, USA: ACM, 2651–2656. DOI:https://doi.org/10.1145/1240866.1241057

E.E. Smith, J. Jonides, and R.A. Koeppe. 1996. Dissociating verbal and spatial working memory using PET. *Cerebral Cortex* 6, 1 (February 1996), 11–20.

Mark A. Staal. 2004. *Stress, Cognition, and Human Performance: A Literature Review and Conceptual Framework*, NASA Ames Research Center.

Robert F. Stanners, Michelle Coulter, Allen W. Sweet, and Philip Murphy. 1979. The pupillary response as an indicator of arousal and cognition. *Motiv Emot* 3, 4 (December 1979), 319–340. DOI:https://doi.org/10.1007/BF00994048

David L. Strayer, Joel M. Cooper, Jonna Turrill, James Coleman, Nate Medeiros-Ward, and Francesco Biondi. 2013. *Measuring Cognitive Distraction in the Automobile*, Washington, DC: AAA Foundation for Traffic Safety.

J. Sweller, J. Merrienboer, and F. Paas. 1998. Cognitive Architecture and Instructional Design. *Educational Psychology Review* 10, 3 (1998), 251–296.

John Sweller. 1994. Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction* 4, 4 (1994), 295–312. DOI:https://doi.org/10.1016/0959-4752(94)90003-5

John Sweller, Paul Ayres, and Slava Kalyuga. 2011. *Cognitive Load Theory*, Springer New York.

J.A. Veltman and A.W. Gaillard. 1998. Physiological workload reactions to increasing levels of task difficulty. *Ergonomics* 41, 5 (May 1998), 656–669. DOI:https://doi.org/10.1080/001401398186829

Julius Verrel, Martin Lövdén, Michael Schellenbach, Sabine Schaefer, and Ulman Lindenberger. 2009. Interacting effects of cognitive load and adult age on the regularity of whole-body motion during treadmill walking. *Psychol Aging* 24, 1 (March 2009), 75–81. DOI:https://doi.org/10.1037/a0014272

Weihong Wang, Zhidong Li, Yang Wang, and Fang Chen. 2013. Indexing Cognitive Workload Based on Pupillary Response Under Luminance and Emotional Changes. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces*. IUI '13. 247–256.

Robert Whelan. 2007. Neuroimaging of cognitive load in instructional multimedia. *Educational Research Review* 2, 1 (2007), 1–12.

Walter W. Wierwille. 1979. Physiological Measures of Aircrew Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 21, 5 (October 1979), 575–593. DOI:https://doi.org/10.1177/001872087902100504

G.F. Wilson and C.A. Russell. 2003. Real-time Assessment of Mental Workload using Psychological Measures and Artificial Neural Network. *Human Factors* 45, 4 (2003), 635–643.

G.F. Wilson and R.E. Schlegel. 2004. *Operator Functional State Assessment*, Neuilly sur Seine, France: :NATO Research and Technology Organization.

Glenn F. Wilson. 2002. An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. *The International Journal of Aviation Psychology* 12, 1 (January 2002), 3–18. DOI:https://doi.org/10.1207/S15327108IJAP1201_2

J. Xu, Y. Wang, F. Chen, and E. Choi. 2011. Pupillary response based cognitive workload measurement under luminance changes. *Human-Computer Interaction–INTERACT 2011* (2011), 178–185.

Bo Yin, Fang Chen, N. Ruiz, and E. Ambikairajah. 2008. Speech-based cognitive load monitoring system. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008*. 2041–2044. DOI:https://doi.org/10.1109/ICASSP.2008.4518041

Galit Yogev, Meir Plotnik, Chava Peretz, Nir Giladi, and Jeffrey M. Hausdorff. 2007. Gait asymmetry in patients with Parkinson's disease and elderly fallers: when does the bilateral coordination of gait require attention? *Experimental Brain Research* 177, 3 (March 2007), 336–346.

K. Yu, J. Epps, and F. Chen. 2011. Cognitive Load Evaluation of Handwriting Using Stroke-level Features. In *Proceedings of the International Conference on Intelligent User Interfaces (IUI 2011)*. Palo, Alto, USA.

Jianlong Zhou, Ju Young Jung, and Fang Chen. 2015. Dynamic Workload Adjustments in Human-Machine Systems Based on GSR Features. In Julio Abascal, Simone Barbosa, Mirko Fetter, Tom Gross, Philippe Palanque, & Marco Winckler, eds. *Human-Computer Interaction – INTERACT 2015*. Lecture Notes in Computer Science. Springer International Publishing, 550–558. DOI:https://doi.org/10.1007/978-3-319-22701-6_40