International Journal of Industrial Ergonomics 50 (2015) 52–61

Contents lists available at ScienceDirect

International Journal of Industrial Ergonomics

journal homepage: www.elsevier.com/locate/ergon

Psychophysically determined user acceptable oral reading speed (UAORS) for an 8-h work day

Tycho K. Fredericks [a], Anil R. Kumar [b, *], Nirari M. Oda [a], Steven E. Butt [a]

[a] Western Michigan University, Kalamazoo, MI 49008, USA
[b] Department of Industrial and Systems Engineering, San Jose State University, San Jose, CA 95192, USA

A R T I C L E I N F O

Article history:
Received 31 October 2014
Received in revised form
3 August 2015
Accepted 4 September 2015
Available online 27 September 2015

Keywords:
Reading speed
Acceptable speech rate
Speech input
Vocal fatigue
Psychophysical approach

A B S T R A C T

Speech input as a form of hands-free input modality has gained popularity over the last decade. At present, there are professions (e.g., medical and legal professionals) who are using this technology (continuous speech) for extended periods of time. Prolonged use of voice-enabled applications might result to voice fatigue and potentially vocal nodules. However, studies providing guidelines on acceptable oral reading speed for an 8-h duration are not available. An experiment using a psychophysical methodology was conducted to determine the user acceptable oral reading speed (UAORS) for an 8-h day. Testing conducted on 10 males indicated that mean speed of 121 words per min could be considered an acceptable oral reading speed for an 8-h day with a 40 dB background noise. A period of 2 h and 20 min of oral reading was also found to be adequate to exert same vocal load as 8 h. Results from this study could be incorporated to set thresholds by designers of speech software in order to prevent vocal fatigue. The acceptable oral reading speed derived in this study could also be used as a guideline for professionals who use their vocal chords for extended periods of time such as teachers, lawyers, clergy, and cheerleaders.

1. Introduction

Speech input as a form of hands-free input modality has gained popularity over the last decade, especially in situations where tasks performed are hands-busy, eyes-busy or performed on the go (i.e. mobility is involved) (Martin, 1989; Stedmon et al., 2003; Al-Qatab and Ainon, 2010). Typically, the vocal inputs/commands are used for speech-to-text, control applications or as a biometrics layer of security. From an algorithmic development perspective, the technology can be segregated as voice recognition and speech recognition. Voice recognition includes the identification of a specific talker's voice from among all other possible and is often used as a security measure. On the other hand, speech recognition is the more general term for the translation of spoken word into text-based representation, which can be used literally (e.g., for dictation) or interpreted as commands to control applications. The speech technology based programs when used as a speech-to-text engine has several perceived advantages such as increased rate of data entry, improved spelling accuracy, permit remote access to databases utilizing wireless technology (Rodger, 2002).

Speech technology was initially targeted for individuals with disabilities since it enabled users to manipulate the machine verbally without having to manually control it and has been employed with some success to control wheelchair (Fezari and Bousbia-Salah, 2007) and switching domestic appliances on and off (Noyes and Frankish, 1992). Rapid advancement due to development of signal processing, algorithms, architectures, and hardware has enabled incorporation of speech recognition based technology in consumer electronics, internet appliances, telephones, automobiles, interactive toys, and industrial, medical, and home electronics and appliances Examples include voice activated services for navigation, and controlling non-essential functions (e.g., radio functions, phone dialing) on cars, voice activated functions on microwaves, interactive voice activated customer services/automated call centers for public utility services, speech enabled search engines such as Google, voice recognition software's for computers, hands free dialing on smart phones, and "Siri" – the popular mobile phone virtual assistant on iPhones. These examples involve voice based services using discrete word or continuous speech for personal or recreational use; however, there are professions (e.g., medical and legal professionals) who are using this

* Corresponding author.
E-mail address: anil.kumar@sjsu.edu (A.R. Kumar).

http://dx.doi.org/10.1016/j.ergon.2015.09.003
0169-8141/

technology (continuous speech) for extended periods of time. Currently, in the medical and legal profession, medical reports and legal transcripts, respectively, are dictated and then transcribed. The transcription is a non-value added step in addition to being expensive and time consuming. Hence, speech recognition used directly by the user (e.g., physician), in conjunction with EHR or as a background process using server-based recognition, becomes a viable option. Additionally, the retrieval of information would also be easier and accessible than the present paper and pen method.

The list of occupations/professionals who are current users or may utilize the technology as part of their routine work in the future is increasing. Provision of voice recognition software on the new Windows and Mac based OS may encourage routine users and professional writers to avail this feature and thereby extend the working duration. Additionally, voice activated gaming systems (e.g., Xbox) and video games (e.g., EndWar, wherein players can issue orders to their troops using their voice), are also available on the market. Abe (2012) developed a hands-free speech input system to record the inspection data while meat inspectors are performing their cattle inspection in a slaughter house. The recorded data is subsequently transferred to a speech recognizer for archiving and quick information retrieval.

The technology also has military applications such as the US Army's Voice Activated Military Tracking Application (VAMTA) (Rodger and George, 2010) and US Navy's Naval Voice Interactive Device (NVID) projects (Simon, 2007). The VAMTA was initially developed as a tool for a paperless method of documentation for diagnostic and prognostic results, leading to automation of maintenance supply actions (Rodger and George, 2010). Recent efforts include adding Siri like national language interface and integrating the system on the Department of Defense (DoD) cloud (Rodger and George, 2010). The NVID utilizes speech recognition systems to enter shipboard environmental survey data with the objective of developing a paperless method for documentation, and reducing the time needed to complete routine inspections such as clinical activities and preventive medicine surveillance, thus giving the corpsmen more time to devote to the primary task of health care responsibilities (Paper et al., 2004).

While speech recognition technology-enabled systems/devices possess extraordinary growth potential, several challenges such as dialectal variations, robustness in noisy environment, and recognition error have impeded the acceptance of this technology when used either as speech to text or command interpretation. A significant technical challenge is the recognition error, especially high when user utilizes a free style of speaking, which results in a negative user experience and eventual frustration. A prime example of which is “Siri” on iPhones. However, based on the advancements in the technology, it may be possible to resolve these technical challenges and have a workable solution in the near future, which potentially implies greater acceptance of this technology by users for extended periods of time. While some users may be the aforementioned population (i.e. people with disability, healthcare and military professionals, writers, or PC users), it is likely that ease of use of the technology may spur users in other professionals such as aircraft pilots, and meat packing inspectors.

With the premise that the speech technology will eventually catch up and free speech without the technical limitations of current speech recognition systems will be available in the near future, it is to be anticipated that users may be required to use more with their voice during their work day. However, prolonged use of this technology will increase the load imposed on the voice which may lead to occupational voice disorders (Kambeyanda et al., 1997; Cudd et al., 1998; Vilkman, 2000). Thus it becomes pertinent to: a) understand the implications of load imposed on the voice due to prolonged time duration, and b) establish guidelines for reading speed during a work shift (8-h workday).

In this preliminary study, we focus on specifically answering “what is an acceptable oral reading speed for an 8 h period when technology is 100%?” The acceptable oral reading speed refers to a user selected oral reading speed (measured in words per minute) that can be sustained without experiencing excessive fatigue over the typical 8-h work duration (hereby referred to as UAORS). The following two sections provide an overview of the pertinent literature related to vocal disorders, and current guidelines for vocal fatigue and oral reading speed over an 8-h workday (either general use or specific to use of speech technology), respectively.

1.1. Vocal disorders

Voice disorders are ubiquitous, and many have severe social, psychological, professional, and economic consequences. It has been suggested that individuals whose professions are vocally demanding may be at greater risk of developing voice disorders than the general population, (Gotaa and Starr, 1993; Williams, 2003). Voice disorders are caused by disturbances in the functioning of the larynx, and can be associated with a variety of conditions, including inflammatory and infectious diseases (Sander and Ripich, 1983), neuroplastic including benign and malignant lesions (Ballantyne and Groves, 1982), psychogenic (Williams and Stevens, 1972), neurological disorders (e.g., nerve damage, Parkinson's disease), neuromuscular (Ludlow et al., 1983), and hyperfunctional voice disorders (Greene and Mathieson, 1989). Hyperfunctional disorders, which includes any behaviors that result in excessive muscular tension in the vocal tract (classified as misuse, overuse, or abuse) is thus of particular interest in the context of this study.

Vocal misuse represents the most common etiologic factor identified in patients with voice disorders (Stemple et al., 1994). Amorim et al. (2011), in a questionnaire designed to define the vocal profile of telemarketing operators, listed 14 complaints related to voice disorders: hoarseness, constant throat clearing, aphonia, sore throat when speaking, fatigue when speaking, sore neck or nape of the neck, blazing when speaking, constant cough, weak voice, voice failure, dry throat, breathiness, soreness when swallowing and the sensation of something stuck in the throat. Morrison et al. (1986) reported that the chief complaint of vocal misuse is muscular fatigue, or a tired and aching throat which increases during the day but recovers with rest. Other factors are excessive talking, overuse, abuse, and excessive prolonged loudness. Excessive talking may lead to degradation in voice quality (slightly rough or hoarse voice), the voice may sound weak, and the person may report that talking requires effort. Symptoms of overuse may include sore throat, neck muscle pain, hoarse and breathy voice, normal pitch may be lowered, loudness cannot be increased, voice starts to cut out at shorter intervals, vocal cords bowing, extreme fatigue in speaking, and difficulty in speaking (Kambeyanda et al., 1997).

It is to be noted that there is a fine distinction between misuse and abuse and the possibility of misuse becoming abuse. An example of abuse is excessive prolonged loudness wherein a person has a habituated pattern of excessively loud voice use and spends much time talking about environmental noise levels. Holbrook (1977) reported that the amount of loud talking was more related to vocal fatigue than the amount of talking during a specified period of time.

Vocal fatigue is a risk for developing speech pathology, including vocal folds lesions, nodules and polyps. The pathologies most common in the younger age group (22–44 years) were vocal nodules and edema (Heaver, 1959; Kleinsasser, 1968), polyps (Wilson, 1987) during middle age (45–64 years), and vocal fold