Int. J. Human–Computer Studies 98 (2017) 1–13

Contents lists available at ScienceDirect

Int. J. Human–Computer Studies

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

Using acoustic paralinguistic information to assess the interaction quality in speech-based systems for elderly users

Humberto Pérez-Espinosa[a,b,*], Juan Martínez-Miranda[a,b], Ismael Espinosa-Curiel[b], Josefina Rodríguez-Jacobo[b], Himer Avila-George[a,b]

[a] *CONACyT, Mexico*
[b] *CICESE-UT[3], Andador 10 #109, Ciudad del Conocimiento, Tepic, Nayarit, Mexico*

ARTICLE INFO

Keywords:
Assistive technology
Elderly user
Interactive system
Paralinguistic phenomenon
Quality of interaction

ABSTRACT

At present, there is a growing need for assisting living systems for the elderly. This type of technology must meet certain accessibility and usability requirements to be adopted successfully. Speech-based interfaces represent a very attractive option for improving the usability and acceptability of assistive technologies for elderly users because they provide a very easily accessible interface and also facilitate paralinguistic analysis of the user's voice to provide valuable information about their emotional and mental states while interacting with a computer system. This information can be used to improve the quality of the interaction between the assistive system and its user. In this study, we propose a method based on the automatic recognition of acoustic paralinguistic phenomena for estimating the quality of speech-based interactions. The proposed method automatically recognizes paralinguistic phenomena (e.g., shouting, hyper-articulation, and hesitation) during the interaction between the user and the system. The interaction is then characterized based on the occurrence of these phenomena. A model is trained using this characterization, which can estimate the quality of the interaction. Using this method, we obtained good recognition performance (F-measure around 70%) when classifying paralinguistic phenomena and there was also a strong correlation between the estimated quality of the interaction metrics and the quality of the interaction metrics reported by the users of an assistive system. The proposed method is useful for automatically estimating the quality of interactions perceived by users and it could support the construction of speech-based systems to adapt and personalize content as well as the style of the interaction with the user.

1. Introduction

During recent decades, the proportion of the population aged over 60 has increased in many countries throughout the world. According to the United Nations, only 8% of the world's population were aged 60 years or over in 1950. However, the proportion had risen to 12% in 2013, and it is expected to reach 21% (more than 2 billion people) in 2050 (D. o. E., 2013). This increase is presenting new challenges in terms of how we can improve the quality of life for elderly people, alleviate pressure on formal care services, and facilitate the tasks of their caregivers (e.g., family members and close relatives). To address these challenges, the research community has proposed various ambient assisted living (AAL) applications for elderly users.

AAL can help older people to execute daily activities in their home and usual environment. AAL may include monitoring behavioral patterns and managing daily activities in the home (Ullberg et al., 2014), promoting social interaction (Bisiani et al., 2013), mobility support (Angeletou et al., 2013), preventing and managing chronic conditions related to age (Tabak et al., 2013), and identifying unsafe situations in the home (Moshtaghi et al., 2015). Most of these applications require continuous and long-term interactions with users to collect relevant data and provide useful feedback. To increase the user acceptance of these types of technology and to fully exploit the potential of AAL applications, it is necessary to achieve more natural interactions. Therefore, it is crucial that the design of interactive user interfaces considers the sensory, motor, and cognitive changes that occur during the aging process. Speech and language technologies can facilitate straightforward interactions, thereby helping to tackle the communication problems experienced by users with chronic conditions and fine-motor problems. There is evidence that among elderly users, speech-based interaction is considered less complicated than traditional forms of interaction, such as using a mouse or keyboard (Taveira

* Corresponding author at: CICESE-UT[3], Andador 10 #109, Ciudad del Conocimiento, Tepic, Nayarit, Mexico.
E-mail addresses: hperez@cicese.mx (H. Pérez-Espinosa), jmiranda@cicese.mx (J. Martínez-Miranda), ecuriel@cicese.mx (I. Espinosa-Curiel), jacobo@cicese.mx (J. Rodríguez-Jacobo), himerag@cicese.mx (H. Avila-George).

http://dx.doi.org/10.1016/j.ijhcs.2016.09.013
Received 12 March 2016; Received in revised form 24 September 2016; Accepted 27 September 2016
Available online 29 September 2016
1071-5819/

and Choi, 2009). The potential of using speech to provide a unified way of interacting with different services and devices was also acknowledged among a set of older adults (Schlögl et al., 2014). However, speech-based interactive systems should also consider the evolution of interactions and the problems associated with speech-based communication, particularly in older adults.

Computational paralinguistics (Schuller and Batliner, 2014) is an emerging research area that involves analyzing the expressions of users to obtain information about individual characteristics (e.g. gender, age, ethnicity, and education level) and different mental states (e.g. confusion, uncertainty, and frustration). Some of the paralinguistic phenomena studied in recent years include the recognition of emotions, attitudes, personality traits, likability, pathology, and the cognitive and physical loads. In addition, the analysis of non-verbal cues in a conversation, such as laughter, fillers, back-channel, silence and overlapping speech, can provide useful information about the gender, role, topic of conversation, mode of interaction, personality, and conflict-handling style of speakers (Vinciarelli et al., 2015a).

The present study had two main aims: (i) to analyze whether there are differences between young and older adults in terms of acoustic paralinguistic information, and automatically assessing the interaction quality by using the paralinguistic phenomena extracted from these two population groups; and (ii) to explore the use of paralinguistic phenomena for automatically assessing the quality of the interaction between users and speech-based systems. The results of these analyses may help to design speech-based interactive systems that can adapt the content and style of interactions with the user by estimating the quality of the interactions. The personalized and continuous adaptation of the contents and style of interaction is a key feature for improving the acceptability of human-computer interaction (HCI) based systems among the target users (Bresó et al., 2016). Moreover, the automatic recognition of paralinguistic phenomena during interactions may help to create better models of human behavior in HCI (Vinciarelli et al., 2015b), e.g., by maintaining an updated model of the users and their context, as well as identifying common characteristics in a group of users (i.e., the older adult population).

The results of this study are expected to be beneficial for AAL applications but also for telecare and personal health systems by complementing the services offered, e.g., in nursing homes and medical facilities. The remainder of this paper is organized as follows. In Section 2, we give a summary of studies that identified differences between young and older adults during the use of general interactive systems, and specifically in speech-based systems. We also describe previous efforts to evaluate the quality of the interaction and user satisfaction when using speech-based systems. In Section 3, we explain the research method, research questions, and experimental design. In Section 4.1, we describe the speech-based interaction corpora used in this study. In Section 4.2, we explain the method used for the automatic identification of a set of acoustic paralinguistic phenomena in elderly users. In Section 4.3, we present the three models used for identifying acoustic paralinguistic phenomena to assess the quality of the interaction, and the results. In Section 5, we discuss the experimental results. In Section 6, we present our discussion, conclusions, and suggestions for future research.

2. Related work

2.1. *Differences between younger and older adults during the use of computer-based interactive systems*

Due to the growth of the older population throughout the world and the high impact of information and communications technologies (ICT) in everyday life, many studies have aimed to understand the requirements for ICT-based systems to facilitate their use and adoption among the older population. Most of these studies involved a comparative analysis between older and younger adults to assess whether there are significant differences in their attitudes toward computer-based systems as well as their abilities (Broady et al., 2010). The results obtained from these studies are not conclusive. Specific studies have reported that the experiences and attitudes of older people toward computers are negative (particularly compared with younger adults) (Laguna and Babcock, 1997; Timmermann, 1998), whereas others studies have indicated that older adults have more positive attitude toward computers than younger adults (Dyck and Smither, 1994). One study concluded that older people generally have a positive attitude toward computers and the use of technology (Eisma et al., 2004).

These contradictory results demand that we focus on the other characteristics of these two user groups rather than simply age-related factors. Some studies have suggested that the primary cause for the negative attitudes toward computers in older adults is their level of experience in the use of these systems (Hawthorn, 2007). According to Charness et al. (2001), older novice learners make more errors and take longer to train than young novice learners. When younger and older adults are experienced users, no age effects are evident, and the number of errors appears to be uniform in both age groups. Older experienced users perform tasks more slowly than younger experienced users, but the difference is smaller than that found in novice users. These differences between younger and older learners in the use of computer-based systems may be closely related to cognitive decay in the elderly (Broady et al., 2010). Cognitive decay in older adults mostly affects reasoning (Horn, 1982) and/or working memory (Salthouse et al., 1991), which are highly related to the process of learning new concepts or skills.

Cognitive decay is not the only factor that influences the performance and attitudes of older adults toward computers, and thus other age-related problems, such as vision, speech and hearing, psychomotor abilities, attention, and automated responses, should be considered in the design of interactive devices and interfaces (Hawthorn, 2000). A mouse and keyboard have been employed most frequently as input devices to assess differences in performance between older and younger adults. Thus, the use of a mouse was evaluated with different actions (click, double click, drag and drop, and menu selection), which showed that older adults made more sub-movements, required longer, and made more errors than younger adults when selecting targets on the screen (Smith et al., 1999; Vigouroux et al.). The main problems for older adults when using a keyboard include confusion with the key labels when users are not familiar with typewriters (Czaja and Lee, 2002), fatigue and discomfort in the forearm and hands during text entry tasks (Piper et al., 2010), and effects on hand functions (Rempel et al., 1999). Other studies have evaluated the use of touch screens by older adults (Wood et al., 2005; Motti et al., 2013), which suggested that a touch screen can accommodate cognitive and hand-eye coordination demands, but the physical demands may make this input device inappropriate for activities that require continuous or ongoing contact (Wood et al., 2005).

The design and development of speech-based systems must also consider the differences found between younger and older adults in previous studies. According to Balota and Duchek (1988), speech slows with age due to the insertion of more and longer pauses into the spoken material and word lengthening. Compared with younger adults, elderly people speak less fluently with more evidence of language planning deficits such as false starts, hesitations and filled pauses, or word repetitions (Kemper, 1992). These specific features of the speech of older adults have been studied in the context of automatic speech recognition (ASR) systems to assess whether the speech of older speakers differs from that of younger speakers in the acoustic feature space (Vipperla et al., 2012), where the aim was to build hierarchical acoustic models using a speech corpus based on age and gender to determine whether improvements in the ASR word error rates were possible in an older age group. The results obtained indicated that gender-dependent models achieved significant improvements compared with speaker-independent models, but only modest additional