

Elderly Speech-Gaze Interaction

State of the Art and Challenges for Interaction Design

Cengiz Acartürk¹(✉), João Freitas², Mehmetcal Fal¹, and Miguel Sales Dias^{2,3}

¹ Informatics Institute, Middle East Technical University, Ankara, Turkey
{acarturk, mehmetcan.fal}@metu.edu.tr

² Microsoft Language Development Center, Lisbon, Portugal
{t-joaof, miguel.dias}@microsoft.com

³ Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR-IUL, Lisbon, Portugal

Abstract. Elderly people face problems when using current forms of Human-Computer Interaction (HCI). Developing novel and natural methods of interaction would facilitate resolving some of those issues. We propose that HCI can be improved by combining communication modalities, in particular, speech and gaze, in various ways. This study presents *elderly speech-gaze interaction* as a novel method in HCI, a review of literature for its potential of use, and discusses possible domains of application for further empirical investigations.

Keywords: Multimodal · Gaze · Eye tracking · Speech · Elderly · Interaction

1 Introduction

A significant advance in human history was the invention of reading: The invention that the visual system could be employed for representing speech. The research in the past several decades has shown that human visual system is active in speech communication in various other ways, such as lip reading and silent speech (thus leading to the term *audio-visual speech*). Human Computer Interaction (HCI) has a broader potential to employ these modalities than it seems at a first glance. One example is the improvement of Automatic Speech Recognition (ASR) by modeling contextual gaze behavior during interaction. We propose that further improvement of HCI can be achieved by designing interfaces for elderly speech and elderly gaze. In the following section, we present a review of these two research fields to provide an interdisciplinary HCI background for *elderly speech-gaze interaction*.

2 Impact of Aging in Communication and Interaction

Elderly individuals have developed resistance to conventional forms of HCI, like the keyboard and mouse, therefore making it necessary to test new natural forms of interaction such as speech, silent speech, touch, gestures, body and head movements, gaze and emotions [1, 2]. In addition, elderly people often have difficulties with motor skills

due to health problems such as arthritis. Therefore, small and difficult-to-handle equipment such as smartphones, may not be easily adopted. It is also known that due to aging, the sensory systems, such as vision become less accurate, therefore difficulties in perception of details or important information in graphical interfaces may arise. On the other hand, current mainstream interfaces, most notably in the mobility area, are rarely designed by taking into account those difficulties that elderly users may face. As a response to those challenges, several devices have been specifically designed or adapted for seniors in the telecommunications market (e.g. Snapfon Ez One, Samsung Jitterbug, ZTC SP45 Senior, etc.).

The broadening of the age-group coverage in user interfaces is necessary given that the population is ageing rapidly in many countries throughout the world, notably, in Europe and Japan. The European Commission estimates that by 2050 the elderly population in the EU (European Union) will be around 29 % of the population. Accordingly, it is hastily becoming necessary to create solutions that allow overcoming age-related difficulties in HCI.

Elderly people who are connected to the world through the internet are less likely to become depressed and have greater probability of becoming socially integrated [3]. Therefore, the internet and user interfaces that allow access to the internet are means for people who want to remain socially active and integrated. In the recent state of technology, however, technological and interaction barriers still do not allow seniors to take a full advantage of the available services and content [1, 4, 5] despite that elderly population is the one that has been more rapidly going online [6].

Several research initiatives and supporting frameworks, have been paving the way to close this gap, with Ambient Assisted Living (AAL) solutions for home and mobility scenarios that have been positively evaluated with elderly populations [1]. We conceive speech systems as a potential complementary solution for HCI usable by elderly speakers, a group of users which has been found to prefer speech interfaces in the mentioned scenarios [1, 7], but also facing limitations in their use due to the inability of these systems to accurately model this population group.

2.1 Elderly Speech

The research literature on elderly speech characteristics does not provide a consistent, general picture. The major source of the divergence is that aging increases the difference between biological age and chronological age, whereas biological aging can also be influenced by factors such as abuse or overuse of the vocal folds, smoking, alcohol consumption, psychological stress/tension, or frequent loud/shouted speech production without vocal training [8, 9]. Accordingly, it may be difficult to determine an exact age limit for *elderly* speech. A usual assumption is that the ages between 60 and 70 are the minimum for the elderly age group [10]. Putting aside the difficulties in the operational definition of elderly age range, there exist specific levels of characterization that make explicit the differences between elderly speech and teenagers or adults speech, such as the acoustic phonetic level [11]. With increasing age there is a deprivation of chest voice, general changes in frequencies, in the voice quality and the timbres. Changes in the heights of vowel formant frequencies particularly occur in older men, not only for

biological reasons, but also because of social changes. Moreover, a slower speech-rate, greater use of pauses, elimination of articles and possessive pronouns, and lower volume of speech are detectable [11–13].

Although being a stable characteristic when compared with the awareness and emotional state of a speaker, human age influences the acoustic signal and the performance of an ASR (Automatic Speech Recognition) engine, as several parameters of the speech wave form are modified, such as fundamental frequency, first and second formants [14], jitter, shimmer and harmonic noise ratio [15]. Those differences between elderly and other user population influence the performance of human-computer interfaces based on speech [16, 17]. This is because the majority of the methods employed for ASR are data-driven. Most techniques (such as Hidden-Markov Models or Deep Neural Networks) model the problem by establishing a generalization that allows inferring recognition results of unseen data. However, different speech patterns such as the ones seen in elderly and children, which are not often used to train such models, cause a decrease in performance of such data-driven systems. The typical strategy to improve ASR performance under these cases is to collect speech data from elderly speakers in the specific domain of the target application and train elderly-only or adapted acoustic models [18–20]. Recent initiatives from the research community that have followed this strategy, specifically for European Portuguese, French, Polish and Hungarian, that targeted speech data collection and acoustic modelling towards the improvement of elderly speech technologies in these language, can be found in the literature [4, 21, 22].

In summary, conventional ASR interfaces do not handle well elderly speech in the recent state of technology. The ASR systems, if trained with young adults' speech, perform significantly worse when used by the elderly population, due to the various mentioned factors as stated above, as well as in the relevant research literature [10]. A solution is to train the systems with elderly speech. However, considerable cost and effort are required for these collections [22]. A complementary solution is to employ modalities other than speech to support ASR. In the present study, we propose that gaze offers this potential to support elderly ASR. The following section presents a brief review of gaze characteristics in elder adults. We believe that this provides the necessary background for using gaze to support elderly speech interfaces.

2.2 Elderly Gaze

Elderly population exhibits different gaze-behavior characteristics than both younger adults and children in some respects. Compared to research on gaze characteristics of younger adults and children, much less research has been conducted for elderly. In this section, our goal is to present an overview of elderly gaze, in comparison to gaze in both younger adults and children where applicable.

A general finding is the loss of inhibitory processing capacity by aging [23], as measured by the so-called antisaccade task. The antisaccade task has been conceived as a measure of general inhibitory control, specifically the control over gaze behavior. In this task, the participant is asked to suppress the reflexive saccade at a visual target that suddenly appears at the periphery of the acute visual field by performing a saccade to the opposite direction of (i.e., by looking away) of the visual target. In particular, the

saccadic reaction time (the time to onset of the eye movement) has been shown to be negatively influenced by age. It has been also shown that elderly participants exhibit longer duration saccades compared to both younger adults and children [24]. This loss in top-down inhibition in elderly may also be observed in patients diagnosed with certain neurological and/or psychiatric disorders [25]. The antisaccade task has also been conceived as an index of cognitive processes, in particular working memory [26], thus as a potential indicator of earlier stages of cognitive decline.

On the other hand, the research literature on aging has revealed that higher-level cognitive abilities are less influenced by aging compared to sensory abilities. For instance, it has been shown that visuospatial memory is not heavily influenced by aging [23]. Similarly, elderly exhibit similar characteristics to younger adults in visual search for targets that are defined by a conjunction of features (in contrast to children, who exhibit slower performance in this task). The major difference between elderly and younger adults is that the elderly have difficulty in moving attention from one item to another [27]. This difference is usually attributed to elderly participants' difficulty in locating peripheral targets rather than a difference in the attentional system between elderly and younger adults [28]. The difficulty in locating peripheral targets is due to a more general finding about the shrinkage in the *useful field of view* (UFoV, the area from which useful information can be extracted) by age [29].

The challenges that elderly people face in reading processes also seem to be related to the shrinkage in the useful field of view (UFoV). Reading comprises both *foveal processing* (for acute visual processing in recognition of letters and words) and *parafoveal processing* (for detecting spaces between words, paragraphs, as well as a few characters to the right of fixation) [30–33]. The previous research on elderly reading shows that elderly readers have a smaller visual span than younger readers [34]. Masking the *foveal* region by means of gaze-contingent eye trackers (thus asking the participants to read *parafoveally*) results in a higher difficulty in elderly readers compared to younger ones [35]. Moreover, a more symmetric visual span is observed in elderly readers. The span in younger adults is asymmetric towards the right or left of fixation, depending on the writing direction. In cultures with left-to-right writing, the span extends to the right of the region, and vice versa for the right-to-left writing cultures. In both cases, the challenge for elderly readers is *parafoveal* processing.

In the studies that present more complex stimuli than the stimuli of the antisaccade task, such as a traffic scene image, the findings address a broader range of eye movement parameters. For instance, in a driving simulation study, the results revealed that elderly participants had more frequent fixations with shorter saccadic amplitudes compared to younger participants. In terms of the scene viewing characteristics, elderly participants spent more time on local regions where younger participants more evenly distributed their gaze throughout the scene [36, 37], accompanied by decreases in elderly drivers' peripheral detection [38]. A similar “tunnel effect (or perceptual narrowing)” phenomenon is observed in elderly drivers in simulated driving context with increased complexity, such as passing maneuvers [39].

These findings suggest that gaze-based interfaces can be used as an interaction method for the elderly. The decrease in inhibitory control of eye movement and the shrinkage in useful field of view (UFoV) by age indicate that a gaze-aware

(i.e., gaze-contingent) interaction has the potential to facilitate visual search and browsing by elderly by providing explicit instructions (e.g., attention attractors) towards the periphery of the visual scene (such as, arrows and other graphical cues that show the direction of the relevant region of interest on the screen).

Finally, there are further aspects of eye movement characteristics that we have not touched upon in the above review. One is pupil size and dilation, which may be employed for detecting emotional states of the participants, as well as cognitive processing difficulties. The general finding is that a smaller maximum dilation velocity characterizes elderly gaze. Moreover, the resting pupil diameter is smaller in elderly compared to younger adults [40]. Recent studies reveal that pupil size is also influenced by processing difficulties in word recognition and response selection in elderly people with hearing loss [41]. Given that elderly users exhibit different emotional patterns, such as the tendency to favor positive stimuli over negative stimuli [42], elderly dilation may require a different interpretation than younger adults, as a measure of users' emotional state. Further research is necessary to reveal the potential of these gaze behavior characteristics in HCI. In the following section, we focus on specific methods of multimodal interaction, which aim at improving elderly speech recognition by gaze.

3 Combining Eye-Gaze Information with Speech

Speech communication is usually a multimodal process in the sense that multiple sources of information are used by humans and affect the way we interpret and issue speech messages. For example, evidence that speech perception employs both hearing and visual senses has been shown by McGurk [43] in 1976. In literature we also find studies that show the use of contextual information such as head and full body movements, gesture, emotions, facial expressions prosody and gaze in human-human speech communication [44–46]. In this analysis we focus our attention in the advantages and disadvantages of the combined use of eye-gaze information along with speech for HCI, since the current literature suggests that a combined application of ASR and gaze information can be used to improve multimodal HCI.

In 2008, Cooke and Russell [47] used the gaze information to change model probabilities of a given word based on the visual focus. In this study the authors assumed a relation between eye movements and the communicative intent. Later studies from the same authors towards noise robust ASR, suggest a relationship between gaze and speech in noisy environments, a “gaze-Lombard effect” [48]. Also in 2008, Prasov and Chai [49] examined the relation between eye-gaze and domain modeling, in a framework that combined speech and eye-gaze for reference resolution. Their conclusions show that eye-gaze information can be used to compensate the lack of domain modeling for reference resolution.

Other authors of multimodal HCI have suggested that the use of gaze information in web-browsing scenarios might provide substantial improvements [50]. This was later verified by Slaney et al. [51]. Slaney et al. reported improvements in ASR performance when accomplishing common browsing tasks such as making a dinner reservation, online shopping and purchasing shoes, and reading online news. The authors used

eye-gaze as contextual information in order to constrain the ASR language model. In terms of the results, improvements of 25 % and 10 % word-error rate (WER) were achieved over a generic and scenario-specific language models. A similar study was conducted by Hakkani-Tür et al. [46] where a conversational web system was developed for interpreting user intentions based on speech and eye-gaze. In this study improvements were also reported not only in predicting the user intention but also in resolving ambiguity, a common technical problem in dialog systems.

Gaze information has also been found to be useful for spoken message dictation scenarios [52–54]. In these studies gaze information has been used as a secondary modality to help choosing between recognition hypotheses in a text entry interface. Additionally, gaze is also used to support correction of speech recognition errors. In the adopted interface model gaze partially replaces the use of the mouse in navigation functions, such as zooming through the presented recognition hypotheses, and for selecting the correct word.

Recent studies also reveal that estimation of eye gaze based on facial pose can have a positive impact in ASR [55]. Other related studies include the analysis of tonal and segmental information, in languages such as Mandarin Chinese, [56], the study of perceptual learning application in speech perception [57], the analysis of the interpolation of lexical, eye gaze and pointing models, which was performed in order to understand aspects of situated dialogues [58], and an in-car multimodal system, which uses information from geo-location, speech, and dialog history, alongside gaze information (estimated from face direction) to interact with the driver [59].

4 Discussion and Conclusion

The studies reviewed in the present paper reveal that eye-gaze information has the potential to significantly increase performance of ASR when combined with speech. However, it is not clear if this fact is applicable to persons of all age groups such as children and elderly.

Analyzing the modalities (speech and gaze) in isolation, and starting by speech, the literature findings suggest that current speech interfaces suffer from its generic modeling approach, not specifically targeted for adult users. Taking into account speech for different age patterns, would resolve this issue but with high cost and effort. As for the literature on eye-gaze information, the studies suggest that it is possible to collect gaze data from all age groups. There is also a fast paced evolution of eye-tracking devices. The cost of desktop-mounted eye tracking sensors has significantly decreased in the past few years, approximately by one hundredth. Although these eye trackers are not appropriate for testing saccade metrics that require high accuracy, they exhibit acceptable performance for spatial precision and accuracy, thus for fixation detection [60]. Therefore it seems plausible to consider its use for real-world scenarios. However, more studies are necessary to have a better understanding of the impact of aging problems upon interaction through eye tracking.

The studies also suggest advantages in the combined used of these modalities. For example, they can be collected in a non-invasive and non-obtrusive manner

(not considering mounted/wearable eye trackers) and allow for a natural interaction with the computer or device. Thus, we believe that elderly could eventually benefit from a multimodal interface based on the analyzed modalities. However, empirical investigations are needed to understand whether a multimodal approach using eye tracking and speech recognition would in fact introduce benefits in HCI with elderly users. Recent research on eye movements and aging reveal that laboratory studies only partly resemble the studies in the real world [61]. Therefore field studies are necessary for testing the settings that are proposed in the present study. In particular, usability studies of multimodal HCI scenarios based on speech and gaze gain relevance and can provide useful feedback about the application of this sort of setups in real environments.

Future work will focus on conducting experimental investigations of the proposed speech-gaze interfaces. This includes exploring novel scenarios such as: (1) the use of gaze combined with speech in mobile scenarios (e.g. interaction with a tablet), which take advantage of new technological solutions in terms of eye tracking; (2) to assess the usability of such multimodal interfaces with users from different age groups, particularly elderly users; (3) to extend the number of HCI tasks which benefit from the combined use of speech and gaze, such as access to online content or interaction with assistive technologies; (4) to understand which of these scenarios can be tackled with an ubiquitous and affordable solution.

Acknowledgment. This work was partially funded by Marie Curie Actions IRIS (ref. 610986, FP7-PEOPLE-2013-IAPP) and METU Scientific Research Project scheme BAP-08-11-2012-121 Investigation of Cognitive Processes in Multimodal Communication.

References

1. Dias, M.S., Pires, C.G., Pinto, F.M., Teixeira, V.D., Freitas, J.: Multimodal user interfaces to improve social integration of elderly and mobility impaired. *Stud. Heal. Technol. Informatics*. **177**, 14–25 (2012)
2. Phang, C.W., Sutanto, J., Kankanhalli, A., Li, Y., Tan, B.C.Y., Teo, H.-H.: Senior citizens' acceptance of information systems: a study in the context of e-government services. *IEEE Trans. Eng. Manag.* **53**, 555–569 (2006)
3. Cisek, E., Triche, K.: Depression and social support among older adult computer users. In: 113th Annual Convention of the American Psychological Association (2005)
4. Oliveira, C., Albuquerque, L., Hämäläinen, A., Pinto, F.M., Dias, M.S., Júdice, A., Freitas, J., Pires, C., Teixeira, V., Calado, A., Braga, D., Teixeira, A.: *Tecnologias de Fala para Pessoas Idosas*. Laboratório Vivo de Usabilidade (Living Usability Lab), pp. 167–181. ARC Publishing (2013)
5. Stephanidis, C., Akoumianakis, D., Sfyarakis, M., Paramythis, A.: Universal accessibility in HCI: process-oriented design guidelines and tool requirements. In: *Proceedings of the 4th ERCIM Workshop on User Interfaces for all*, Stockholm, pp. 19–21 (1998)
6. Fox, S.: Are “wired seniors” sitting ducks? Pew Internet & American Life Project (2006)
7. Teixeira, V., Pires, C., Pinto, F., Freitas, J., Dias, M.S., Rodrigues, E.M.: Towards elderly social integration using a multimodal human-computer interface. In: *Proceedings of International Living Usability Lab Workshop on AAL Latest Solutions, Trends and Applications (AAL 2012)*, pp. 3–13 (2012)

8. Jessen, M.: Speaker classification in forensic phonetics and acoustics. In: Müller, C. (ed.) *Speaker Classification 2007*. LNCS (LNAI), vol. 4343. Springer, Heidelberg (2007)
9. Linville, S.E.: *Vocal Aging*. Singular Thomson Learning, CA (2001)
10. Wilpon, J.G., Jacobsen, C.N.: A study of speech recognition for children and the elderly. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 1996)*, pp. 349–352. IEEE (1996)
11. Helfrich, H.: *Age Markers in Speech*. Cambridge University Press, Cambridge (1979)
12. Pellegrini, T., Hämäläinen, A., de Mareüil, P.B., Tjalve, M., Trancoso, I., Candeias, S., Dias, M.S., Braga, D.: A corpus-based study of elderly and young speakers of european portuguese: acoustic correlates and their impact on speech recognition performance. In: *Proceedings of (INTERSPEECH 2013)*, pp. 852–856 (2013)
13. Stover, S.E., Haynes, W.O.: Topic manipulation and cohesive adequacy in conversations of normal adults between the ages of 30 and 90. *Clin. Linguist. Phon.* **3**, 137–149 (1989)
14. Albuquerque, L., Oliveira, O., Teixeira, T., Sá-Couto, P., Freitas, J., Dias, M.S.: Impact of age in the production of european portuguese vowels. In: *15th Annual Conference of the International Speech Communication Association (INTERSPEECH 2014)*, Singapore (2014)
15. Xue, S.A., Hao, G.J.: Changes in the human vocal tract due to aging and the acoustic correlates of speech production a pilot study. *J. Speech Lang. Hear. Res.* **46**, 689–701 (2003)
16. Pellegrini, T., Trancoso, I., Hämäläinen, A., Calado, A., Dias, M.S., Braga, D.: Impact of age in asr for the elderly: preliminary experiments in European Portuguese. In: Torre Toledano, D., Ortega Giménez, A., Teixeira, A., González Rodríguez, J., Hernández Gómez, L., San Segundo Hernández, R., Ramos Castro, D. (eds.) *IberSPEECH 2012*. CCIS, vol. 328, pp. 139–147. Springer, Heidelberg (2012)
17. Schultz, T.: Speaker characteristics. In: Müller, C. (ed.) *Speaker Classification 2007*. LNCS (LNAI), vol. 4343, pp. 47–74. Springer, Heidelberg (2007)
18. Anderson, S., Liberman, N., Bernstein, E., Foster, S., Cate, E., Levin, B., Hudson, R.: Recognition of elderly speech and voice-driven document retrieval. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1999)*, pp. 145–148. IEEE (1999)
19. Baba, A., Yoshizawa, S., Yamada, M., Lee, A., Shikano, K.: Elderly acoustic model for large vocabulary continuous speech recognition. *IEICE Trans. Inf. Syst.* **J85-D-2(3)**, 390–397 (2002)
20. Vipperla, R., Wolters, M., Georgila, K., Renals, S.: Speech input from older users in smart environments: challenges and perspectives. In: Stephanidis, C. (ed.) *UAHCI 2009, Part II*. LNCS, vol. 5615, pp. 117–126. Springer, Heidelberg (2009)
21. Hämäläinen, A., Avelar, J., Rodrigues, S. Dias, M., Kolesinski, A., Fegyó, T., Németh, G., Csobánka, P., Lan, K., Hewson, D.: The EASR corpora of European Portuguese, French, Hungarian and polish elderly speech. In: *LREC*, pp. 1458–1464 (2014)
22. Hämäläinen, A., Pinto, F., Dias, M., Júdice, A., Freitas, J., Pires, C., Teixeira, V., Calado, A., Braga, D.: The first European Portuguese elderly speech corpus. In: *Proceedings of IberSPEECH, Madrid, Spain (2012)*
23. Olincy, A., Ross, R.G., Youngd, D.A., Freedman, R.: Age diminishes performance on an antisaccade eye movement task. *Neurobiol. Aging* **18**, 483–489 (1997)
24. Munoz, D.P., Broughton, J.R., Goldring, J.E., Armstrong, I.T.: Age-related performance of human subjects on saccadic eye movement tasks. *Exp. Brain Res.* **121**, 391–400 (1998)
25. Munoz, D.P., Everling, S.: Look away: the anti-saccade task and the voluntary control of eye movement. *Nat. Rev. Neurosci.* **5**, 218–228 (2004)

26. Bowling, A., Draper, A.: Using saccadic eye movements to assess cognitive decline with ageing. In: Horsley, M., Eliot, M., Knight, B.A., Reilly, R. (eds.) *Current Trends in Eye Tracking Research*, pp. 237–244. Springer International Publishing, Switzerland (2014)
27. Trick, L.M., Enns, J.T.: Lifespan changes in attention: the visual search task. *Cogn. Dev.* **13**, 369–386 (1998)
28. Scialfa, C.T., Thomas, D.M., Joffe, K.M.: Age differences in the useful field of view: an eye movement analysis. *Optom. Vis. Sci. Official Publ. Am. Acad. Optom.* **71**, 736–742 (1994)
29. Beurskens, R., Bock, O.: Age-related decline of peripheral visual processing: the role of eye movements. *Exp. Brain Res.* **217**, 117–124 (2012)
30. Kliegl, R., Nuthmann, A., Engbert, R.: Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *J. Exp. Psychol. Gen.* **135**, 12–35 (2006)
31. Rayner, K.: Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* **124**, 372–422 (1998)
32. Rayner, K., Pollatsek, A., Ashby, J., Clifton, C.: *The Psychology of Reading*, 2nd edn. Psychology Press, Abingdon (2012)
33. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Halszka, J., van de Weijer, J.: *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press, Oxford (2011)
34. Rayner, K., Castelano, M.S., Yang, J.: Eye movements and the perceptual span in older and younger readers. *Psychol. Aging* **24**, 755–760 (2009)
35. Rayner, K., Yang, J., Schuett, S., Slattery, T.J.: The effect of foveal and parafoveal masks on the eye movements of older and younger readers. *Psychol. Aging* **29**, 205–212 (2014)
36. Maltz, M., Shinar, D.: Eye movements of younger and older drivers. *Hum. Factors* **4**, 15–25 (1999)
37. Ho, G., Scialfa, C.T., Caird, J.K., Graw, T.: Visual search for traffic signs: the effects of clutter, luminance, and aging. *Hum. Factors* **432**, 194–207 (2001)
38. Schieber, F., Gilland, J.: Age differences in the useful field of view during real-world driving. In: *Proceedings of the Human Factors and Ergonomic Society Annual Meeting*, vol. 49, pp. 182–185. Sage Publications (2005)
39. Cantin, M.L.M.T.V., Teasdale, M.S.N.: Aging yields a smaller number of fixations and a reduced gaze amplitude when driving in a simulator. *Adv. Transp. Stud. Int. J. Special Issue*, 21–30 (2006)
40. Bitsios, P., Prettyman, R., Szabadi, E.: Changes in autonomic function with age: a study of pupillary kinetics in healthy young and old people. *Age Ageing* **25**, 432–438 (1996)
41. Kuchinsky, S.E., Ahlstrom, J.B., Vaden, K.I., Cute, S.L., Humes, L.E., Dubno, J.R., Eckert, M.A.: Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology* **50**, 23–34 (2013)
42. Mather, M., Carstensen, L.L.: Aging and motivated cognition: the positivity effect in attention and memory. *Trends Cogn. Sci.* **9**, 496–502 (2005)
43. McGurk, H., MacDonald, J.: Hearing lips and seeing voices. *Nature* **264**, 746–748 (1976)
44. Oviatt, S.: Ten myths of multimodal interaction. *Commun. ACM* **42**, 74–81 (1999)
45. Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X.-F., Kirbas, C., McCullough, K.E., Ansari, R.: Multimodal human discourse: gesture and speech. *ACM Trans. Comput. Interact.* **9**, 171–193 (2002)
46. Hakkani-Tür, D., Slaney, M., Celikyilmaz, A., Heck, L.: Eye gaze for spoken language understanding in multi-modal conversational interactions. In: *Proceedings of the 16th International Conference on Multimodal Interaction*, pp. 263–266. ACM (2014)
47. Cooke, N.J., Russell, M.: Gaze-contingent automatic speech recognition. *Signal Process. IET* **2**, 369–380 (2008)

48. Cooke, N., Shen, A., Russell, M.: Exploiting a “gaze-Lombard effect” to improve ASR performance in acoustically noisy settings. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014), pp. 1754–1758. IEEE (2014)
49. Prasov, Z., Chai, J.Y.: What’s in a gaze? the role of eye-gaze in reference resolution in multimodal conversational interfaces. In: Proceedings of the 13th International Conference on Intelligent User Interfaces, pp. 20–29. ACM (2008)
50. Heck, L.P., Hakkani-Tür, D., Chinthakunta, M., Tür, G., Iyer, R., Parthasarathy, P., Stifelman, L., Shriberg, E., Fidler, A.: Multi-modal conversational search and browse. In: SLAM Workshop, pp. 96–101 (2013)
51. Slaney, M., Rajan, R., Stolcke, A., Parthasarathy, P.: Gaze-enhanced speech recognition. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014), pp. 3236–3240. IEEE (2014)
52. Vertanen, K.: Efficient correction interfaces for speech recognition (2009)
53. Vertanen, K., MacKay, D.J.C.: Speech dasher: fast writing using speech and gaze. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 595–598. ACM (2010)
54. Vertanen, K., MacKay, D.J.C.: Speech dasher: a demonstration of text input using speech and approximate pointing. In: Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility, pp. 353–354. ACM (2014)
55. Slaney, M., Stolcke, A., Hakkani-Tür, D.: The relation of eye gaze and face pose: Potential impact on speech recognition. In: Proceedings of the 16th International Conference on Multimodal Interaction, pp. 144–147. ACM (2014)
56. Malins, J.G., Joannis, M.F.: The roles of tonal and segmental information in Mandarin spoken word recognition: an eyetracking study. *J. Mem. Lang.* **62**, 407–420 (2010)
57. Mitterer, H., Reinisch, E.: No delays in application of perceptual learning in speech recognition: evidence from eye tracking. *J. Mem. Lang.* **69**, 527–545 (2013)
58. Kennington, C., Kousidis, S., Schlangen, D.: Interpreting situated dialogue utterances: an update model that uses speech, gaze, and gesture information. In: Proceedings of the SIGDIAL 2013 (2013)
59. Misu, T., Raux, A., Lane, I., Devassy, J., Gupta, R.: Situated multi-modal dialog system in vehicles. In: Proceedings of the 6th Workshop on Eye Gaze in Intelligent Human Machine Interaction: Gaze in Multimodal Interaction, pp. 25–28. ACM (2013)
60. Dalmaijer, E.S.: Is the low-cost EyeTribe eye tracker any good for research? PeerJ PrePrints (Preprint, 2015). <http://dx.doi.org/10.7287/peerj.preprints.585v1>, <https://peerj.com/preprints/585/>, Accessed 25 Jan 2015
61. Dowiasch, S., Marx, S., Einhauser, W., Bremmer, F.: Effects of aging on eye movements in the real world. *Frontiers Hum. Neurosci.* **9**, 46 (2015)