Can older people remember medication reminders presented using synthetic speech?

RECEIVED 26 March 2014
REVISED 26 June 2014
ACCEPTED 3 July 2014
PUBLISHED ONLINE FIRST 30 July 2014

Maria K Wolters[1], Christine Johnson[2], Pauline E Campbell[3], Christine G DePlacido[2], Brian McKinstry[4,5]

ABSTRACT

Reminders are often part of interventions to help older people adhere to complicated medication regimes. Computer-generated (synthetic) speech is ideal for tailoring reminders to different medication regimes. Since synthetic speech may be less intelligible than human speech, in particular under difficult listening conditions, we assessed how well older people can recall synthetic speech reminders for medications. 44 participants aged 50–80 with no cognitive impairment recalled reminders for one or four medications after a short distraction. We varied background noise, speech quality, and message design. Reminders were presented using a human voice and two synthetic voices. Data were analyzed using generalized linear mixed models. Reminder recall was satisfactory if reminders were restricted to one familiar medication, regardless of the voice used. Repeating medication names supported recall of lists of medications. We conclude that spoken reminders should build on familiar information and be integrated with other adherence support measures.

Key words: Medication Adherence, Synthetic Speech, Older People, Interface Design, Reminders

BRIEF COMMUNICATION

BACKGROUND AND SIGNIFICANCE

As more people live longer, levels of multimorbidity have increased, leading to more complex medication regimes. Non-adherence can have serious consequences, including hospitalization and death.[1,2] Forgetting what to take when is a common cause of non-adherence.[3] Reminders that help users remember those details are an important part of successful adherence interventions.[4–6]

Spoken reminder messages can be presented through many channels such as interactive voice response systems, a digital TV, or electronic pill boxes. Using a computer-generated (synthetic) voice, spoken messages can be tailored quickly and cost effectively to different patients and medication regimes, which may make them more effective at changing medication behavior.[7]

However, despite major advances in speech synthesis technology, synthetic speech is still not as intelligible as human speech,[8] and the perceived listening effort is significantly higher.[9] In this study, we assess whether synthetic speech is sufficiently intelligible to allow older people (who are likely to have hearing loss) to recall medication names, which are difficult to remember.[10]

Hearing loss starts early. About one in four adults aged 50–59 has a hearing loss in the frequencies covered by speech sounds,[11] and incidence rises sharply with age.[12] As more cognitive and perceptual effort is required to process auditory stimuli, fewer cognitive resources may be available for understanding and remembering what was said.[13]

Older people find synthetic speech more difficult to understand than younger people.[14,15] However, most related work[14–19] has focused on older speech synthesis technology and may therefore overestimate intelligibility problems. Here, we used publicly available implementations of the two main current speech synthesis approaches, statistical parametric synthesis[20,21] and unit selection[22,23] (figure 1). Both methods start with a large database of speech. Given a message, unit selection systems search the database for pieces of speech that fit the message and concatenate them. Statistical parametric systems use the database to build a set of statistical models of all core speech units, which are then used to generate arbitrary new messages.

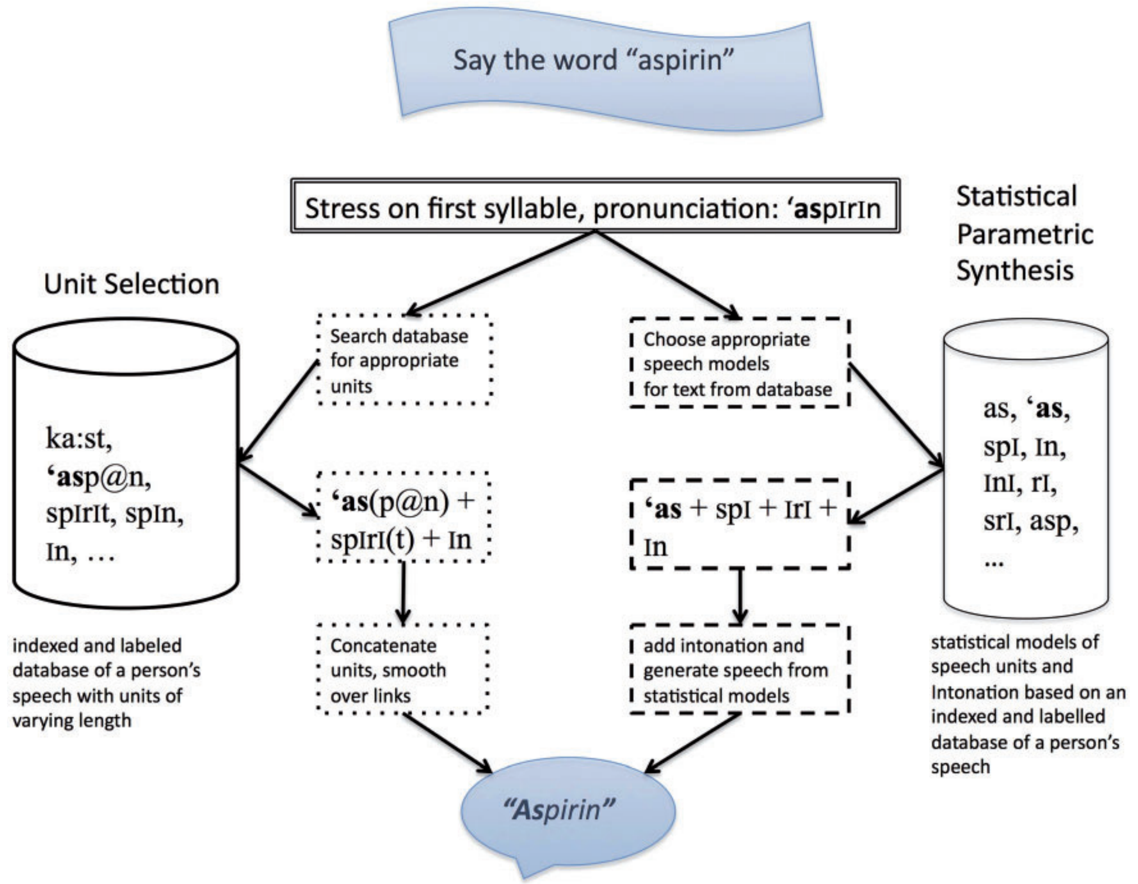
Reminders are often heard against background noise, such as radio, TV, or road traffic, or they can be transmitted through noisy telephone lines, which may make them more difficult to understand. Listeners may also be distracted by other tasks or people after hearing a reminder. Therefore, we varied listening conditions in our study design and included a brief distraction before recall.

Better message phrasing and organization can counteract unfavorable listening conditions. In our study, we tested two options: adapting the message to how listeners think about

Correspondence to Dr Maria K Wolters, School of Informatics, University of Edinburgh, Informatics Forum IF 3.46, 10 Crichton Street, Edinburgh EH8 9AB, UK; maria.wolters@ed.ac.uk

For numbered affiliations see end of article.

Figure 1: One word, two ways of converting it to speech: unit selection (left) and statistical parametric synthesis (right).



medication, for example by linking each medication to the reason for taking it,[24,25] and repeating key information.[26]

OBJECTIVES

This study was designed to determine if older people with a range of hearing ability (from clinically normal to some age-related hearing loss) find medication reminders more difficult to recall when these are presented using a synthetic as opposed to a human voice. We included two factors that might impair recall: memory load (one vs four medications per reminder) and listening conditions (difficult vs acceptable). Finally, we tested whether message design might be able to counteract recall problems. Two alternatives were tested: explaining medication indications (Explanation) and repeating medication names (Repetitions).

MATERIALS AND METHODS

Stimuli

Based on 12 frequently taken UK medications (six over-the-counter and six prescription), we created 12 one-medication reminders, 24 baseline four-medication reminders, 12 four-medication reminders where the medications were repeated, and 12 four-medication reminders where indications were explained (table 1).

Auditory stimuli were generated using three male voices with a standard British English accent: a human voice, a statistical parametric synthesis voice,[21] and a unit selection voice.[23]

Participants heard stimuli in two levels of background noise (multi-speaker speech shaped babble,[27] signal-to-noise ratios 0 and +10) and two levels of signal transmission quality (noisy and clear telephone line).

To reduce duration, we only tested the Explanation and Repetition message designs for high-memory load reminders with four medications. Participants heard both designs under two listening conditions, clear phone line/soft background noise (acceptable) and noisy phone line/loud background noise (difficult).

Participants

Ethics approval for this study was granted by the South East Scotland NHS Ethics Board, reference number 10/S1103/43. Participants were recruited from four family practices in

Table 1: Examples of the medication reminders used in the study

Reminder	Example
One medication	
Baseline	Please remember to take the following medication: aspirin.
Four medications	
Baseline	Please remember to take the following four medications: aspirin, Corsodyl, amoxicillin, and Dulcolax
Repetition	Please remember to take the following four medications: aspirin, Corsodyl, amoxicillin, and Dulcolax. I repeat: aspirin, Corsodyl, amoxicillin, and Dulcolax
Explanation	Please remember to take the following four medications: aspirin, to thin your blood, Corsodyl, for your mouth ulcer, amoxicillin, for your infection, and Dulcolax, for your constipation.

Edinburgh through the Scottish Primary Care Research Network. We contacted native speakers of English aged 50+ with no hearing aids, no cognitive impairment, no neurodegenerative disorder, and sufficient mobility to attend testing. Of 56 participants who attended, 44 met the inclusion criteria.

Procedure

In a questionnaire, participants indicated whether they were familiar with the 12 medications used in the task. For each familiar medication, they were asked to describe its use.

Participants were screened for cognitive impairment using the Addenbrooke's Cognitive Examination-Revised.[28] Working memory capacity was assessed by a reading span test.[29] Working memory scores were recoded into three levels: first quartile, second/third quartile, and fourth quartile.

Hearing levels were measured using pure-tone audiometry,[30,31] and severity was categorized following Martini and Mazzoli.[32] Participants were considered to have a hearing loss if they had at least a mild hearing loss in their poorer ear. We excluded people with a severe hearing loss in at least one ear or a conductive hearing loss (following ref. [33]).

The intelligibility test consisted of a total of eight training and 72 test trials. Participants adjusted loudness to a comfortable level at the start of the experiment.

In each trial, participants heard a medication reminder, followed by a 2 s pause and a simple five-word distractor sentence.[34] After repeating the distractor, participants had to recall the medications in the reminder. Figure 2 shows a sample trial.

To support recall, participants were shown a list of 24 possible medications and their key indication. Twelve targets that occurred in the reminders and 12 alternatives matched for indication. Responses were audio-recorded and transcribed verbatim by the experimenter. The reminder score was the number of correctly remembered medications.

All participants heard the same reminder/distractor pairs in the same, randomized order. The assignment of voices to sentences was balanced using a Latin square design.

Participants were debriefed in a semi-structured exit interview.

Figure 2: Sample intelligibility trial.

Sample Intelligibility Trial

System:	Please remember to take the following four medications: Aspirin, Corsodyl, Amoxicillin, and Piriton. [2 second pause] Rachel wins three old mugs.
Experimenter:	[Switches to sentence screen]
User:	"Rachel wins two old mugs."
Experimenter:	[Types "Rachel wins two old mugs.] [Switches to medication screen which asks user to recall medications and displays 24 medications, 12 target and 12 distractor]
User:	"Aspirin, Piriton, and something else."
Experimenter:	[ticks Aspirin and Piriton on screen, switches to next task]

Statistical analysis

Intelligibility scores were analyzed using generalized linear mixed models (GLMM[35,36]). Following established best practice for GLMMs,[36] the effect of each predictor is illustrated with a graph showing the 95% CI for that predictor. Fisher's exact test and pairwise t tests corrected for multiple comparisons as implemented in R (V.3.0.2) were used to illustrate effects of single predictors.

RESULTS

Demographic and baseline data are summarized in table 2. The decline in working memory capacity and the prevalence of hearing loss in the sample is consistent with normal ageing.[11,37]

On average, participants knew five of the over-the-counter and two of the prescription medications.

Recall was perfect for 74.8% (790 of 1056) of the one-medication reminders, but only for 4.5% (95 of 2112) of the four-medication reminders. When participants made a mistake, they mostly forgot a medication name or confused the medication with another target medication.

Figure 3 shows the results of the GLMM analyses for one medication, four medications (baseline message design only), and four medications (message design varied).

Table 2: Demographics of participants

Gender		
Female	N = 24	55%
Male	N = 20	45%
Age		
Mean (SD), years	64 (9)	
50–59	N = 15	34%
60–69	N = 14	32%
70 +	N = 14	32%
No information	N = 1	2%
Education (highest level achieved)		
Postgraduate	N = 11	25%
Graduate	N = 9	20%
Vocational	N = 6	14%
Secondary school*	N = 14	32%
None	N = 1	2%
No information	N = 3	7%
Hearing loss (on worse ear)		
None	N = 26	59%
Mild, monaural	N = 7	16%
Mild, binaural	N = 8	18%
Moderate on worse ear, mild on better ear	N = 2	4.5%
Moderate, binaural	N = 1	2%
Working memory capacity†		
Mean (SD)	43 (13)	
Age 50–59	51 (12)	
Age 60–69	42 (15)	
Age 70 +	38 (11)	

*Secondary school qualifications include GCSE, Leaving Certificate, and A-Levels.
†The score is the sum of all items recalled correctly in the correct order. Scores range from 0 to 70.

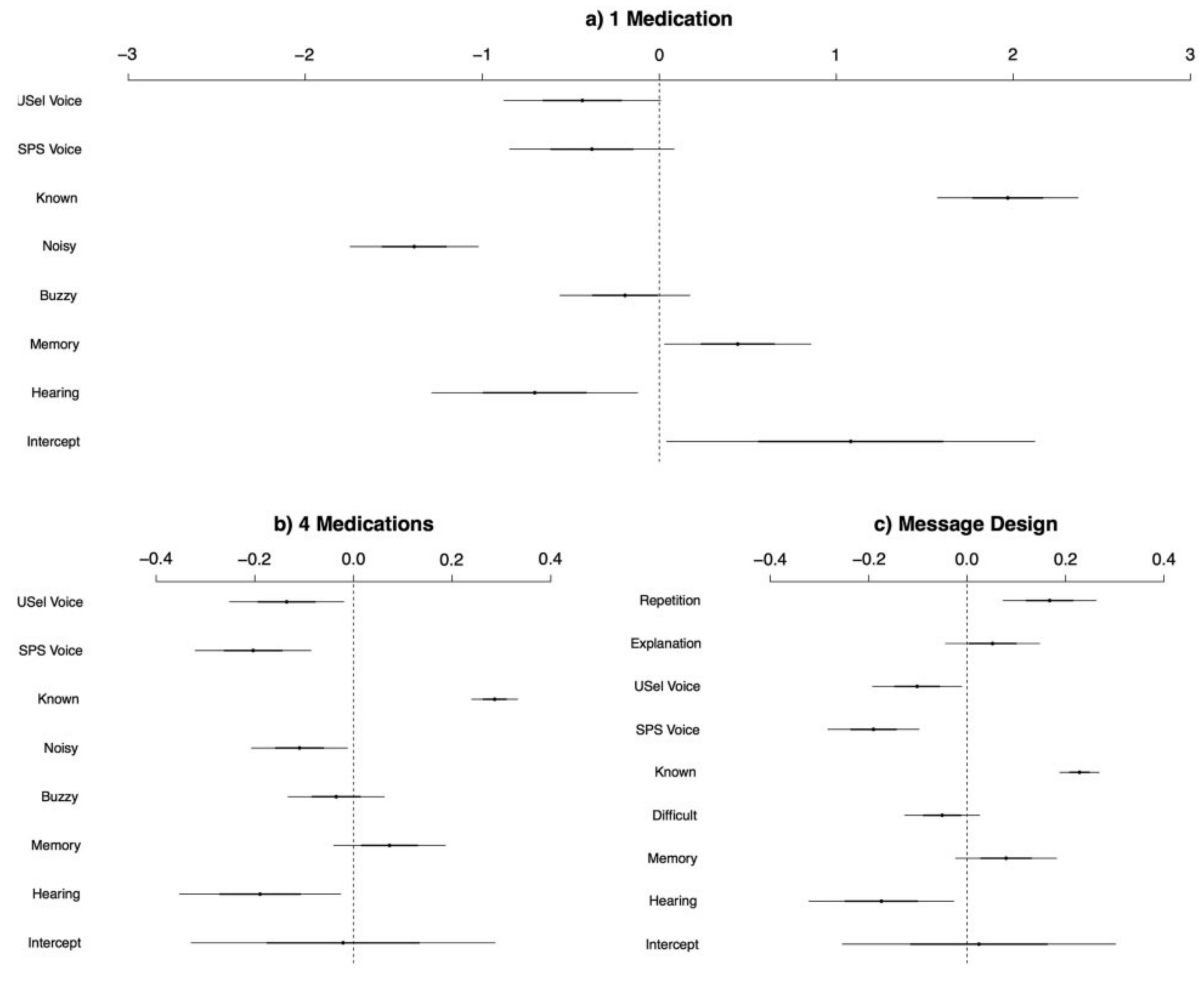
As expected, listeners with a hearing loss found it more difficult to recall the reminders than listeners with normal hearing. Working memory capacity does not have a significant effect after controlling for hearing. Signal quality did not affect recall; the effect of background noise varied depending on the number of medications and message design. Figure 3 illustrates the corresponding parameter values.

Looking at one-medication reminders, we find that known medications are typically recalled correctly (89.3% (484 of 541) recall for known, 59.5% (306 of 512) recall for unknown; $p<0.001$, OR 5.7, 95% CI 4.1 to 8.0).

Although voice type affected recall, this difference was almost entirely due to unknown medications. For known medications, participants recalled 89% (human voice), 90% (statistical parametric synthesis), and 88% (unit selection), while for unknown medications, they recalled 64.8% (human voice) and 52.2% (both synthetic voices).

We found similar patterns in the four-medication case. Listeners knew an average of 1.9 medications and remembered 1.6. The average number of medications recalled for the human voice was 1.86 (SD 1.1), 1.5 (SD 1.1) for statistical

Figure 3: Graphical summary of the three generalized linear mixed models fitted for the main analysis. (A) Model for one medication only. (B) Model for four medications, no variation in message design. (C) Model for four medications, message design varied. The graphs show 95% CIs for the estimates of each individual level predictor. Group level predictors are not shown. In the text, predictors are referred to as having a significant effect if the 95% CI of the corresponding estimate does not include 0. The individual level corresponds to the variation among trials, while the group level covers variation among participants. We used a default intercept and the predictors presence of hearing loss (Hearing), working memory capacity level (Memory), signal quality (Signal), background noise level (Noise), medication familiarity (Known), and voice type (SPS: statistical parametric synthesis, USel: unit selection). For the Message Design model, we added predictors corresponding to message type, Explanation and Repetition. The Signal and Noise conditions were collapsed into one predictor, difficult listening condition (Difficult). On the participant level, we used a default intercept that captured inter-individual variation.



parametric synthesis, and 1.6 (SD 1) for the unit selection voice (figure 3B).

Message design improved recall. With the basic design, listeners recalled an average of 1.6 medications (SD 1.0), with explanations, 1.8 (SD 1.1), and with repetitions, 2 (SD 1.0). Only the effect of repeating medication names was significant after controlling for all other relevant variables (figure 3C).

Repeating medications particularly helped listeners understand the synthetic voices. In the difficult listening condition (loud noise, bad signal), listeners recalled an average of 1.7 medications (SD 1.1) in the human voice, but only 1.4 (SD 1.1) when medications were presented using statistical parametric synthesis, and 1.5 (SD 1.0) when they heard the unit selection voice. Using repetitions, recall increased to a mean of 2.1 (human, SD 1.0), 1.9 (statistical parametric synthesis, SD 1.1), and 2.1 (unit selection, SD 1.1). Unit selection performed as well as the human voice (pairwise comparisons using the t test with adjustments, $p<0.6$) and outperformed statistical parametric synthesis ($p<0.02$).

BRIEF COMMUNICATION

DISCUSSION AND CONCLUSION

When presenting short reminders for known medications, a high-quality synthetic voice works as well as a human voice. However, for longer lists of medication names or unknown medication names, recall is not satisfactory, regardless of voice.

Therefore, spoken medication reminders will only work if they build on what users know. Knowledge can come from users' own experience of taking medications, from friends and family, or from the media. In practice, this means that spoken reminders have to be integrated into complex interventions that allow for patient input.

Speech synthesis method matters. While both synthetic voices performed similarly overall, unit selection, which preserves more acoustic information, was more robust when lists of medication names were repeated, and statistical parametric synthesis performed better for one-medication reminders. Reminders could also be sent directly to users' hearing aids, bypassing much of the ambient distortion.

Repeating key information is significantly better at improving recall than explanations, even though in the exit interviews, participants thought that the explanations were particularly helpful. While repetition alone is not enough to help people recognize or recall information,[38,39] in this case, it provides a second chance to catch an auditory glimpse of the speech signal,[40] building a more robust percept that is in turn more likely to be recognized later.

Limitations

The distractor task was designed to prevent participants from rehearsing the medication name(s) subvocally, a common strategy for remembering auditory information.[41] We need to examine the effect of more realistic distractors that use modalities other than hearing. We also did not probe why participants were familiar with a given medication.

It is not clear why working memory had little effect. This could be due to the high proportion of participants with relatively good working memory. Working memory might also be less important than other cognitive abilities, such as speed of processing.

In line with best practice in designing health information technology,[42] our study was conducted in a laboratory environment. In deployment, reminders will need to be tailored to fit with a person's concepts of their medication and their routine, and to reduce the potential for unintended disclosure of illness.[43,44]

The medication list provided an overview of possible medications similar to a list provided by a clinician, or to a medicine cabinet where people can see all their medications at once. However, real medicine cabinets (and medication lists) are often far more messy.[45,46] Reading level may also have affected participants' ability to use the list.

Finally, we used actual medication names instead of phrases such as 'morning medication' that patients typically use to describe medications. However, such phrases require clinicians and pharmacists to cooperate with patients and carers. Otherwise, patients may not remember what pills are included in the 'morning medication.'

CONCLUSION

In order that the usefulness and usability of computer-generated reminders in practice can be assessed, they need to be integrated into a multimodal medication management system that also provides tailored information about other aspects of the medication such as dosage and side effects.

ACKNOWLEDGEMENTS

We thank Vasilis Karaiskos for programming and conducting the listening tests, Elaine Niven and Dina van der Hulst for training in the administration and scoring of cognitive tests, and Martin Cooke, Elaine Niven, and Catherine Mayo for valuable discussion. Simulations of signal transmission quality were kindly provided by the German Telekom Usability Lab, Technical University of Berlin.

CONTRIBUTORS

MW, BM, and CDP designed the study; MW, CDP, CJ, and PC piloted the design; MW conducted the statistical analysis and prepared a first draft of the paper; and all authors discussed the interpretation of results and read and commented on several drafts of the paper.

FUNDING

This project was funded by the Chief Scientist Office Scotland, grant number CZG/2/494, and the EPSRC/BBSRC initiative SPARC.

COMPETING INTERESTS

None.

ETHICS APPROVAL

South East Scotland NHS Ethics Board approved this study (reference number 10/S1103/43).

PROVENANCE AND PEER REVIEW

Not commissioned; externally peer reviewed.

DATA SHARING STATEMENT

The quantitative data and the R code used for the statistical analysis is available from the first author on request for non-commercial purposes. It cannot be passed on without the first author's consent. This paper should be cited in any publications arising from the data, and the data set may be anonymized further.

REFERENCES

1. Cresswell KM, Fernando B, McKinstry B, *et al*. Adverse drug events in the elderly. *Br Med Bull* 2007;83:259–74.
2. Budnitz DS, Pollock DA, Weidenbach KN, *et al*. National surveillance of emergency department visits for outpatient adverse drug events. *JAMA* 2006;296:1858–66.

3. Barber N. Patients' problems with new medication for chronic conditions. *Qual Saf Heal Care* 2004;13:172–5.
4. Dunbar PJ. A two-way messaging system to enhance antiretroviral adherence. *J Am Med Informatics Assoc* 2002;10:11–15.
5. Russell CL, Conn VS, Jantarakupt P. Older adult medication compliance: integrated review of randomized controlled trials. *Am J Health Behav* 2006;30:636–50.
6. Haynes RB, Ackloo E, Sahota N, *et al*. Interventions for enhancing medication adherence. *Cochrane Database Syst Rev* 2008;(2):CD000011.
7. Revere D, Dunbar PJ. Review of computer-generated outpatient health behavior interventions: clinical encounters "in absentia". *J Am Med Informatics Assoc* 2001;8:62–79.
8. King S, Karaiskos V. The Blizzard challenge 2012. Proceedings of the Blizzard Challenge Workshop; 2012.
9. King S, Karaiskos V. The Blizzard challenge 2013. Proceedings of the Blizzard Challenge Workshop; 2013.
10. Wolters MK, Campbell P, DePlacido C, *et al*. Making synthetic speech accessible to older people. Proceedings of the 6th ISCA Workshop on Speech Synthesis, Bonn, Germany; 2007:288–93.
11. Ciletti L, Flamme GA. Prevalence of hearing impairment by gender and audiometric configuration: results from the National Health and Nutrition Examination Survey (1999–2004) and the Keokuk County Rural Health Study (1994–1998). *J Am Acad Audiol* 2008;19:672–85.
12. Cruickshanks KJ, Wiley TL, Tweed TS, *et al*. Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin. *Am J Epidemiol* 1998;148:879–86.
13. Wingfield A, Tun PA, McCoy SL, *et al*. Sensory and cognitive constraints in comprehension of spoken language in adult aging. *Semin Hear* 2006;27:273–83.
14. Hardee JB, Mayhorn CB. Reexamining synthetic speech: intelligibility and the effects of age, task, and speech type on recall. Human Factors and Ergonomics Society Annual Meeting Proceedings. Human Factors and Ergonomics Society; 2007:1143–7.
15. Roring RW, Hines FG, Charness N. Age differences in identifying words in synthetic speech. *Hum Factors* 2007;49:25–31.
16. Al-Awar Smither J. Short term memory demands in processing synthetic speech by old and young adults. *Behav Inf Technol* 1993;12:330–5.
17. Sutton B, King J, Hux K, *et al*. Younger and older adults' rate performance when listening to synthetic speech. *Augment Altern Commun* 1995;11:147–53.
18. Paris CR, Thomas MH, Gilson RD, *et al*. Linguistic cues and memory for synthetic and natural speech. *Hum Factors* 2000;42:421–31.
19. Dulude L. Automated telephone answering systems and aging. *Behav Inf Technol* 2002;21:171–84.
20. Zen H, Tokuda K, Black AW. Statistical parametric speech synthesis. *Speech Commun* 2009;51:1039–64.
21. Yamagishi J, Watts O. The CSTR/EMIME HTS system for Blizzard challenge 2010. Proceedings of the BLIZZARD Challenge; 2010.
22. Hunt A, Black AW. Unit selection in a concatenative speech synthesis system using a large speech database. ICASSP-96. Vol 1. Atlanta, Georgia, 1996:373–6.
23. Richmond K, Strom V, Clark R, *et al*. Festival multisyn voices for the 2007 Blizzard challenge. Proceedings of the 3rd Blizzard Challenge; 2007.
24. Morrow D, Carver LM, Leirer VO, *et al*. Medication schemas and memory for automated telephone messages. *Hum Factors* 2000;42:523–40.
25. Morrow D, Leirer VO, Carver LM, *et al*. Repetition improves older and younger adult memory for automated appointment messages. *Human Factors* 1999;41:194–204.
26. Morrow DG, Leirer VO, Carver LM, *et al*. Effects of aging, message repetition, and note-taking on memory for health information. *J Gerontol B Psychol Sci Soc Sci* 1999;54:369–79.
27. Dreschler WA, Verschuure H, Ludvigsen C, *et al*. ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. International Collegium for Rehabilitative Audiology. *Audiology* 2001;40:148–57.
28. Mioshi E, Dawson K, Mitchell J, *et al*. The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *Int J Geriatr Psychiatry* 2006;21:1078–85.
29. Niven E. An exploration of representation and maintenance in visuo-spatial working memory for simultaneously and sequentially presented information [Unpublished PhD Thesis]. University of Edinburgh, 2010.
30. British Society of Audiology. British Society of Audiology guidelines on minimum training standards for otoscopy and impression taking. 2004.
31. British Society of Audiology. British Society of Audiology guidelines on pure tone air and bone conduction threshold audiometry with and without masking. 2011.
32. Martini A, Mazzoli M. Achievements of the European Working Group on Genetics of Hearing Impairment. *Int J Pediatr Otorhinolaryngol* 1999;49(Suppl 1):S155–8.
33. British Society of Audiology. British Society of Audiology guidelines on tympanometry. 2013.
34. Wagener KC, Brand T, Kollmeier B. International cross-validation of sentence intelligibility tests. Proceedings of the Conference of the European Federation of Audiological Sciences; 2007:8–10.
35. Brown H, Prescott R. *Applied mixed models in medicine*. 2nd edn. Chichester, UK: John Wiley & Sons, 2006.
36. Gelman A, Hill J. *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press, 2007.
37. Johnson W, Logie RH, Brockmole JR. Working memory tasks differ in factor structure across age cohorts: implications for dedifferentiation. *Intelligence* 2010;38:513–28.

38. Tulving E. How many memory systems are there? *Am Psychol* 1985;40:385–98.
39. Bekerian DA, Baddeley AD. Saturation advertising and the repetition effect. *J Verbal Learning Verbal Behav* 1980;19: 17–25.
40. Cooke M. A glimpsing model of speech perception in noise. *J Acoust Soc Am* 2006;119:1562.
41. Baddeley A. The fractionation of working memory. *Proc Natl Acad Sci USA* 1996;93:13468–72.
42. Klasnja P, Consolvo S, Pratt W. How to evaluate technologies for health behavior change in HCI research. Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems—CHI '11. New York, New York, USA: ACM Press, 2011:3063.
43. Littenberg B, MacLean CD, Hurowitz L. The use of adherence aids by adults with diabetes: a cross-sectional survey. *BMC Fam Pract* 2006;7:1.
44. McGee-Lennon MR, Wolters MK, Brewster S. User-Centred Multimodal Reminders for Assistive Living. CHI '11: Proceedings of the 29th International Conference on Human Factors in Computing Systems; 2011.
45. Siek KA, Khan DU, Ross SE, *et al*. Designing a personal health application for older adults to manage medications: a comprehensive case study. *J Med Syst* 2011;35: 1099–121.
46. Wieczorkiewicz SM, Kassamali Z, Danziger LH. Behind closed doors: medication storage and disposal in the home. *Ann Pharmacother* 2013;47:482–9.

AUTHOR AFFILIATIONS

[1]School of Informatics and School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, Edinburgh, UK

[2]Speech and Hearing Sciences, Queen Margaret University, Edinburgh, UK

[3]NMAHP Research Unit, Glasgow Caledonian University, Glasgow, UK

[4]Centre for Population Health Sciences, University of Edinburgh, Edinburgh, UK

[5]Edinburgh Health Services Research Unit, Edinburgh, UK