



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Hold that thought: are spearcons less disruptive than spoken reminders?

Citation for published version:

Wolters, M, Isaac, K & Doherty, J 2012, Hold that thought: are spearcons less disruptive than spoken reminders? in *CHI '12 Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, pp. 1745-1750, 2012 ACM SIGCHI Conference on Human Factors in Computing Systems, Austin, Texas, United States, 5/05/12. <https://doi.org/10.1145/2212776.2223703>

Digital Object Identifier (DOI):

[10.1145/2212776.2223703](https://doi.org/10.1145/2212776.2223703)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Published In:

CHI '12 Extended Abstracts on Human Factors in Computing Systems

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Hold That Thought: Are Spearcons Less Disruptive than Spoken Reminders?

Maria K. Wolters
School of Informatics,
University of Edinburgh
10 Crichton Street
Edinburgh, EH8 9AB, UK
maria.wolters@ed.ac.uk

Karl B. Isaac
School of Informatics,
University of Edinburgh
10 Crichton Street
Edinburgh, EH8 9AB, UK
k.b.isaac@sms.ed.ac.uk

Jason M. Doherty
School of Informatics,
University of Edinburgh
10 Crichton Street
Edinburgh, EH8 9AB, UK
j.m.doherty@sms.ed.ac.uk

Abstract

Speech reminders can severely disrupt list recall. Spearcons, time-compressed speech messages, might be less disruptive because they are much shorter. In this study, we asked 24 younger participants to recall 64 short lists of digit, animal, food, or furniture names. List items were presented one at a time; the number of items presented depended on individual digit spans. Spearcons affected list recall to the same extent as speech. However, people with higher digit spans had significantly worse recall. This could be due to short-term memory overload or the longer presentation time of long lists. We discuss implications for menu design.

Keywords

working memory; reminders; speech; Spearcon; Irrelevant Speech Effect

ACM Classification Keywords

H.5.1 [Multimodal Interfaces]: Audio Input/Output.;
H.5.2 [User Interfaces]: Natural Language.

General Terms

Human Factors, Experimentation

Copyright is held by the author/owner(s).
CHI 2012, May 5–10, 2012, Austin, TX, USA.
ACM 978-1-4503-1016-1/12/05.

Sue, aged 75, is writing her shopping list in the living room when she hears the message “Hello Sue! In one hour, you need to be at your mammography screening. It’s time to get ready and call the taxi”. Sue had seen the appointment on the diary screen in the kitchen at lunch time, but she was so focused on remembering all the ingredients she needed that she forgot to keep an eye on the time.

Sue’s neighbours often visit for a quick chat. Since Sue doesn’t want any gossip, she has used the “fast” option of her audio reminder system. She can understand the fast voice because she is used to it and knows what to listen out for, but to her friends, it is just garbled.

Introduction

When a reminder, warning, or alarm is played, the intended recipient is usually involved in doing something else. Auditory reminders are particularly useful when the recipient is doing something that requires their visual attention. However, not all auditory messages are created equal. Beeps, which have been used in notification systems for people with memory problems (e.g., [7]), merely alert the user that something else requires attention. While speech can be more specific (e.g., [9]), it has its own problems—spoken messages are more intrusive, less private, and longer than brief sounds [11, 12]. Speech also disrupts the recall of a list of items (Irrelevant Speech Effect, [2]). This effect is stronger for long messages than for short keywords [16]. Non-speech can also evoke this effect if changes in pitch, tempo, and timbre are similar to those of speech [15].

Alternatives to beeps and full spoken messages include Earcons [1], tunes with an arbitrary association between tune and meaning, and auditory icons [4], sounds that represent the meaning of a message in the same way that a visual icon does. Spearcons (e.g., [3]) are a compromise between short non-speech audio stimuli and full speech. They consist of spoken messages that have been highly time-compressed. Since Spearcons are much shorter than speech, this could mitigate the disruption. They are also more difficult to understand for strangers.

Although Spearcons have been studied in dual task situations [8], we are not aware of any work that tests the extent to which Spearcons trigger the Irrelevant Speech Effect. In this pilot study, we tested the effect of spoken reminders and their Spearcon versions on the recall of lists of items of varying length and semantic category. Typically, words for single digits are used. In order to

make the lists more realistic, we also used item sets that consisted of words for animals, food and furniture.

Method

Participants

We recruited 24 participants aged 18–35 (median age: 21.5). All had at least a high school education. Two thirds ($n = 16$) were female. Digit span was established using the Forward Digit Span test from the Wechsler Adult Intelligence Scale, Version IV [17]. All participants passed a hearing screen at 25 dB(HL) on the better ear and completed the physical and health sub-scales of the RAND SF-36 Quality of Life questionnaire [6].

Task and Materials

We used eight words used per list content type. The words used in the animal, food, and furniture lists were high-frequency words with 1–2 syllables; numerical words were monosyllabic digit names. For each type of list item (animals, food, furniture, numbers), we generated random sequences of length DS-1, DS-2, DS-3, and DS-4, where DS is the participant’s forward digit span. Since longer lists are harder to recall, we will refer to this parameter as *difficulty*.

List items were presented one at a time for one second each. They were displayed in the middle of an iPad 2 screen in portrait orientation, black on white, using 240 point Helvetica. Participants saw each combination of list length and content four times, twice with and twice without a reminder.

Speech reminder messages were generated using Festival 2.1 and the HTS 2010 *Roger* voice [18]. Spearcons were generated from these files using software provided by the Georgia Tech Sonification Lab. The average duration of HTS speech messages was 2.1s (SD=0.3s). Spearcons

were on average 35.6% of the original length (SD=3.2%). The mean Spearcon duration was 0.74s (SD=0.04s).

Participants were familiarised with messages and associated commands in a training session. First, for each reminder, they heard the HTS speech message, followed by the Spearcon version. They then read aloud the response, which was shown as soon as the Spearcon had finished playing. In part two, the responses were no longer displayed and had to be recalled by the participant.

Participants practised the list recall task without reminders eight times, twice for each list content type, using two lists of length 3 and 5. This ensured that they had seen all of the eight words of each content type.

Each participant completed 64 tasks, which were presented in a randomised order. In all tasks, responses were verbal and scored by an experimenter. In half the tasks, participants saw a list of items and had to repeat them in the right order after a wait of 15 seconds. In the other half, they heard a reminder message at the start of the 15 second wait and then had to issue the appropriate command. Half the messages were Spearcons, half were HTS speech. Each message occurred twice.

We created four sets of tasks, each with two blocks, using a Latin-square design to ensure that an equal number of participants saw a given combination of list items and reminder content.

Statistical Analysis

For data analysis, we used a mixed logistic regression model. The model had one group-level predictor, participant ID, and seven individual-level predictors: digit span, difficulty, type of list item, reminder (none/HTS speech/Spearcon), digit span \times difficulty, list item type \times

reminder, and list item type \times difficulty. In order to assess the contribution of each individual-level predictor, we added these to the model sequentially. We started with a model that consisted of an intercept and the group-level term. In each subsequent step, we added the predictor that covered the largest amount of residual variation in the data. The impact of each predictor was assessed using an analysis of deviance and the χ^2 test of significance. We used the Akaike Information Criterion (AIC) as a rough measure of model quality.

Results

Participants' mean digit span was 6.8 (SD 1.2, range 5 – 9). Participants scored an average of 100% on the Physical scale of the SF-36, indicating excellent self-reported physical condition (SD 5%, range 75% – 100%). The mean score on the Health subscale was 75% (SD 16.6%, range 15% – 100%).

Table 1: Contribution of sequentially adding individual-level predictors to model. No.: sequence in which predictors were added. Lower AIC = better model.

No.	Predictor	AIC	p (χ^2)
1	Reminder	1667	$p < 0.0001$
2	Difficulty	1451	$p < 0.0001$
3	Digit Span	1424	$p < 0.0001$
4	Type of List Item	1388	$p < 0.0001$
5	Diff. \times Item Type	1379	$p < 0.005$
6	Diff. \times Digit Span	1371	$p < 0.005$
7	Item Type \times Reminder	1370	$p < 0.1$

Participants responded correctly to an average of 96.0% reminders (SD 5.3%, range 87.5% – 100%) HTS speech and Spearcon reminders were recalled equally well (HTS: M=96.9%, SD=4.9; Spearcon: M=95.0%, SD=6.4;

Wilcoxon signed rank test for matched pairs, $p < 0.2$). List recall was substantially worse than reminder recall. On average, 62.7% of all lists were repeated correctly (SD=19.5%, range: 25%–90%). The effects of digit span, list content, difficulty, and reminder type on recall are summarised in Table 1, which shows the sequence in which predictors were added to the baseline model (AIC: 1864.0), and the resulting improvement in the model.

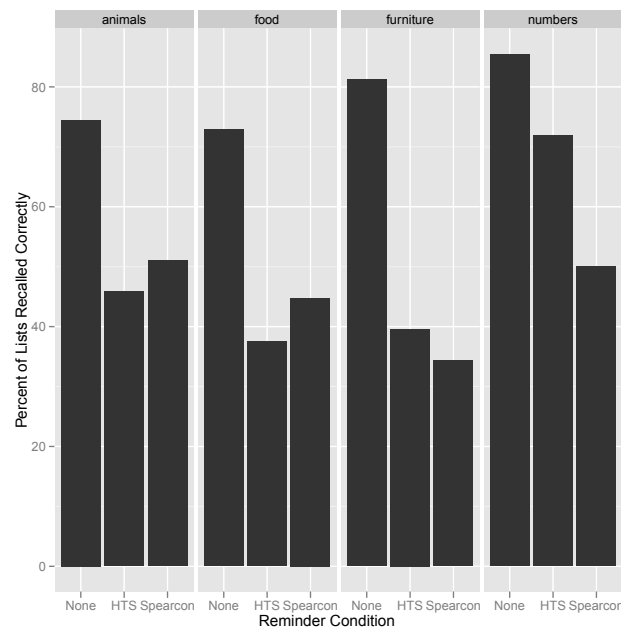


Figure 1: Effect of list content and reminder type on list recall.

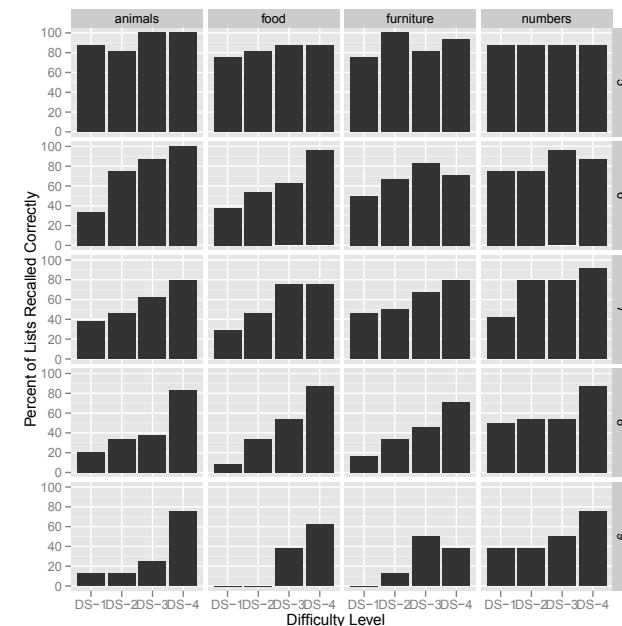


Figure 2: Effect of adjusted list length (difficulty level) on correct list recall. DS-4 is the shortest (easiest), DS-1 is the longest (hardest). Results are grouped by participant digit span (5–9, vertical) and list content (animals, food, furniture, numbers, horizontal)

The most important predictor of list recall is whether a reminder is present. As Fig. 1 shows, both HTS and Spearcon reminders disrupt recall. When no reminder is present, participants recall four in five (78.52%) of all lists correctly. When they need to respond to a reminder, this falls to one in two (HTS speech: 48.70%, Spearcon: 45.05%). There is no difference between both reminder types (Wilcoxon signed rank test, $p < 0.26$). Lists of digit words are recalled best (overall, 73.18% correct), followed

by animal names (61.46%), furniture items (59.11%), and food items (57.03%).

However, more important than list content are two predictors that our design was supposed to control for, digit span and task difficulty. The longer a person's digit span, the more they struggled with longer lists (c.f. Fig. 2), in particular when they consisted of animal, food, or furniture names instead of words for single digits ("numbers").

Discussion and Future Work

Participants responded reliably and appropriately to both Spearcon and HTS speech messages. This is in line with the results of Jeon et al. [8], who used Spearcons and speech for menu navigation. Both types of status message disrupted the serial recall to a similar extent.

There are several reasons why lists with words for digits were recalled better than lists of animal, food, or furniture. Digits are highly familiar, remembering sequences of digits such as phone numbers is a common task, and the original Wechsler Digit Span task also involved lists of digits, so there might be a practice effect.

The decreased performance of participants with high digit span could be explained by time-based decay [14]. Briefly, the memory trace of an item starts to decay after presentation, and the longer recall is delayed, the less likely the item is to be remembered. For a list of length 8, participants had to remember the first item for a total of 15 (delay before recall) + 7 (time for presenting the remaining 7 items) = 22 seconds. If this explanation is correct, we would expect people with higher digit spans to perform much better when they see the whole list at the same time.

Another possible reason is limited memory capacity. According to the multi-component model of working-memory [13], verbal input passes through a short-term store called the phonological loop. It is possible that lists of more than four items overload the phonological loop even for participants with high digit spans. If this is indeed the case, removing the time delay should not improve performance.

Deciding between both explanations is relevant for presenting lists where the user needs to remember each item. If time decay is important, pauses between items should be minimised, if the capacity of the phonological loop is exhausted, lists should be no longer than 4–5 elements.

This experiment was a design pilot for a larger study which will be concerned with older people's perception of Spearcons. Before the main study, we will conduct a second pilot investigating the role of time decay versus memory capacity in remembering lists of items to ensure this confounder will be adequately controlled. Since older people find time-compressed speech more difficult to learn than the younger people who participated in our study [5], we hypothesise that they may also find Spearcons more difficult to process, despite training. The results will feed into a field study of a prototype home reminder system as part of the MultiMemoHome project [10].

Acknowledgements

This research was funded by grant no. EP/G060614/1 from the UK Engineering and Physical Sciences Research Council. We thank our participants for their time, and Ben Davison and Bruce Walker for their generous help with creating Spearcons.

References

- [1] S. A. Brewster. Non-speech auditory output. In J. A. Jacko and A. Sears, editors, *The Human-Computer Interaction Handbook*, pages 220–239. Lawrence Erlbaum, Mahwah, NJ, 2002.
- [2] H. A. Colle and A. Welsh. Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behaviour*, 15:17–31, 1976.
- [3] T. Dingler, J. Lindsay, and B. Walker. Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. In *Proc. ICAD*, 2008.
- [4] W. Gaver. The SonicFinder: An interface that uses auditory icons. *Human-Computer Interaction*, 4:67–94, 1989.
- [5] J. Golomb, J. Peelle, and A. Wingfield. Effects of stimulus variability and adult aging on adaptation to time-compressed speech. *Journal of the Acoustical Society of America*, 121:1701–1708, 2007.
- [6] R. Hays, C. Sherbourne, and R. Mazel. User's Manual for Medical Outcomes Study (MOS) Core Measures of Health-Related Quality of Life. 1995.
- [7] E. Inglis, A. Szymkowiak, P. Gregor, A. F. Newell, N. Hine, B. A. Wilson, J. Evans, and P. Shah. Usable technology? Challenges in designing a memory aid with current electronic devices. *Journal of Neuropsychological Rehabilitation*, 14:77–87, 2004.
- [8] M. Jeon, B. Davison, M. Nees, J. Wilson, and B. Walker. Enhanced auditory menu cues improve dual task performance and are preferred with in-vehicle technologies. In *Proceedings of AutomotiveUI*, page 91, Sept. 2009.
- [9] L. Lines and K. Hone. Multiple voices, multiple choices: Older adults' evaluation of speech output to support independent living. *Gerontechnology*, 5(2):78–91, 2006.
- [10] M. McGee-Lennon, M. Wolters, and S. Brewster. User-Centred Multimodal Reminders for Assistive Living. In *Proceedings of CHI*, 2011.
- [11] M. R. McGee-Lennon, M. K. Wolters, and T. McBryan. Auditory reminders in the home. In *Proceedings of ICAD*, Montreal, 2007.
- [12] J. Noyes, E. Hellier, and J. Edworthy. Speech warnings: A review and future directions. *Theoretical Issues in Ergonomic Science*, 7:551–571, 2006.
- [13] G. Repovs and A. Baddeley. The multi-component model of working memory: explorations in experimental cognitive psychology. *Neuroscience*, 139:5–21, 2006.
- [14] J. N. Towse, G. J. Hitch, and U. Hutton. On the interpretation of working memory span in adults. *Memory cognition*, 28(3):341–8, 2000.
- [15] S. Tremblay, A. Nicholls, D. Alford, and D. Jones. The irrelevant sound effect: Does speech play a special role? *J Exp Psychol: Learning, Memory, and Cognition*, 26:1750–1754, 2000.
- [16] R. Vilimek and T. Hempel. Effects of speech and non-speech sounds on short-term memory and possible implications for in-vehicle use. In *Proc. of ICAD*, 2005.
- [17] D. Wechsler. Wechsler Adult Intelligence Scale IV, 2008.
- [18] J. Yamagishi and O. Watts. The CSTR/EMIME HTS system for Blizzard Challenge 2010. In *Proc. BLIZZARD Challenge*, 2010.