Multiple voices, multiple choices: Older adults' evaluation of speech output to support independent living

Lorna Lines
Kate S. Hone
VIVID Research Centre, School of Information Systems, Computing & Mathematics, Brunel University, Uxbridge, UB8 3PH, United Kingdom
E: Lorna.Lines@Brunel.ac.uk

L. Lines, K.S. Hone, Multiple voices, multiple choices: Older adults' evaluation of speech output to support independent living, Gerontechnology 2006; 5(2):78-91.
Interactive Domestic Alarm Systems (IDAS) are currently being designed to support older adults with independent living. Uptake and continued use of an IDAS will likely depend on successful interactions between the older adult and the system. Thus, suitable modes of system output must be identified to facilitate IDAS-User interactions. This paper considers the use of speech as a mode of IDAS output and reports design recommendations arising from an experimental investigation into older adults' subjective evaluations of speech. Methodological reflections consider the suitability of employing an experimental methodology with an ageing user group.

Keywords: older adults, speech output, interactive domestic alarm systems

Due to transformations in health, social and lifestyle trends the UK's ageing population is rapidly increasing, with 31% of the UK's population predicted to be over the age of 60 by the year 2051[1]. Although a cause for celebration, these figures also present problems. Local authorities across the UK report resource concerns regarding service provision for older adults[2]. With an expected increase in the aged population and many older adults requiring local authority housing and care services, greater demands will be made on already under-resourced service providers. To alleviate pressures on these services, there is currently greater emphasis on the role of technologies to support older adults within their own homes[3-5]. Smart home technology is now being utilised to develop *assistive technologies* such as Interactive Domestic Alarm Systems (IDASs), designed specifically to assist older adults with living independently in their own homes for longer.

This paper introduces IDASs and considers the importance of the design of the user interface for an ageing user group within the domestic context; specifically focussing on the subjective evaluation of speech as a mode of IDAS output. Personal preferences, operational context and system functionality as factors of subjective evaluation influence are recognised. Following this discussion, three critical speech output design factors [speech source (natural, synthetic), speech gender (male, female) and environmental conditions (quiet, noisy)] are identified and subsequently investigated using an experimental methodology. Discussion of the findings of this investigation considers each design factor to inform IDAS design recommendations. Methodological reflections are also presented, together with suggestions for modifications to traditional experimental design when employing an ageing user group.

Interactive domestic alarm systems

To explain how an IDAS operates, take, for example, the Millennium Home System (*Figure 1*), an academic and industrial collaboration based at Brunel University. It operates using sensors, retrofitted to the fabric of the older adult's home that detect changes in the domestic environment. Sensor data is continuously monitored and analysed by a central computer that, upon detection of an alarm state, activates the user interface to alert the user to the alarm state. Alarm states can be considered as any change within the domestic environment that might cause negative, if not fatal consequences (directly or indirectly) for the user, for example, doors and windows left open, taps left running or if the user has had a fall. Having received an alarm state 'alert' such as "The system has detected that the back door is open and needs to be closed", it is expected that the user will, either (i) resolve the alarm state, (ii) cancel the alarm state via manipulation of a dedicated input device, or (iii) request human assistance. If the user resolves the alarm successfully the system will resume a 'safe state'. If the user cancels the alarm the IDAS will continue to monitor the alarm from a 'sleeping' state, reactivating the alarm and re-alerting the user following a pre-determined amount of time. However, if the user requests assistance or if user feedback is undetected, the central computer will initiate an alarm call, via a call centre, to an external care agency requesting human assistance. A 'safe' system state will only resume when human assistance has been received.

Differences between IDASs will lie in the specific technologies utilised, the number and type of alarm states that can be detected and, perhaps most critically, the design of the user interface.

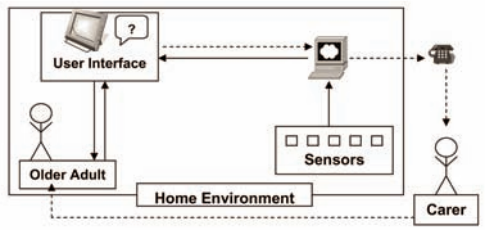

Figure 1. The Millennium Home System

The IDAS user and the user interface

The user interface is often the solitary element upon which interactive systems are evaluated[3]. User interfaces for general systems are typically designed for younger adults[6] and little research exists on the design of user interfaces for older adults[7]. From a human factors perspective, the design and development of an IDAS user interface must consider the diversity in cognitive, physical and sensory abilities presented within an ageing user group. Here, IDAS users can be defined as older adults over the age of 65[1], who are living independently but may need assistance at times due to natural ageing declines in physical or psychological function, with levels of decline differing between individuals. This variability and diversity must be accommodated by the IDAS user interface to facilitate effective IDAS-User interactions.

Current research suggests designing multimodal interfaces to accommodate older adults presenting declines in one or more, cognitive, physical and sensory abilities[8-10]. Multimodal interfaces can employ a variety of input and output device, such as buttons, pens, speech and visual displays to facilitate user-system interaction. Effective IDAS-User interaction is critical for the successful functionality of an IDAS. If the user cannot interact with an IDAS, alarm states may go unresolved and lead to negative, possibly fatal, consequences for the older adult. Here, this investigation takes a first step in considering which,

of the large number of input and output devices available, could be used within an IDAS multimodal user interface; specifically considering the effectiveness of speech output.

Speech as IDAS output

Speech output is utilised successfully within a variety of technological applications in military and industrial environments and within numerous consumer products and services, such as aircraft cockpits, cars and educational children's toys. Speech may therefore be an effective output mode for an IDAS. A potential benefit of speech output arises from the omnidirectional nature of the human auditory system, whereby speech output can have an immediate effect on the auditory system unlike a visual display that must be within the users' viewpoint to be noticed. In addition, a strong association between visual impairment and ageing[11] suggests many older adults might have problems receiving visual outputs but may be able to receive spoken outputs. It is acknowledged that for older adults who present severe hearing losses, speech output may be unsuitable. However, for older adults presenting slight/moderate declines in auditory acuity, simply increasing system volume may overcome reported difficulties. Furthermore, in the domestic environment it is likely that the user may engage in hands/eyes busy tasks, for example, cooking and cleaning. Unlike visual outputs, users can receive speech outputs while conducting a concurrent hands/eyes busy task[12] and these do not require the user to accommodate their location/position to receive alarm state information.

From these benefits, speech appears to be an appropriate mode of IDAS output in relation to the user, the domestic environment and IDAS functionality. Thus it would appear appropriate at this point to consider the alternative types of speech output currently available.

Types of speech output

Existing speech output can be classified as natural (human) speech or synthetic speech. Natural speech outputs are created via digital recordings of human speakers. Although this may retain the naturalness of the speech output, vocabulary is limited by the number of recordings of natural speech and once recorded, changes cannot be made to the vocabulary, intonation or vocal properties[13]. Thus, natural speech outputs are suitable for systems requiring predefined vocabularies that are unlikely to change over time and can be produced in male and female formats dependent on the 'speaker' selection.

In contrast synthetic speech is generated by technologies that simulate human speech. For example, concatenated synthesis presents pre recorded sub-word units of human speech[13] and is often implemented in systems requiring small, but dynamic vocabularies. Once again, this form of speech synthesis can create male and female formats based on 'speaker' selection. This is due to the high number of sub-word unit recordings, the large system storage space and the computational power required for reassembly. It is likely that during the early stages of development IDASs will present limited functionality and will require only a small vocabulary; Therefore neither natural nor concatenated speech output can be ruled out as potential sources of IDAS speech output at this stage of the development.

User evaluation of speech output

Previously, speech output has been shown to achieve a functional level of comprehension within the IDAS context[14,15]. Although it is critical to comprehend IDAS outputs, it is argued here that the user's subjective evaluation of speech output may have an equal impact on the uptake and continued use of an IDAS.

User comprehension does not predict user evaluation and vice versa[16]. Even if the user can understand speech as IDAS output, this does not mean that the user will positively evaluate the spoken information presented; alternatively, if the user evaluates the IDAS speech output negatively this is not indicative of poor comprehension. If an alternative mode is chosen this may be ineffective or difficult for the user to access when compared to the use of speech as IDAS output. To maximise positive user evaluations and therefore optimise the potential for continued use, alternative types of speech output should be considered and the effectiveness of each measured in terms of IDAS user evaluation.

User evaluation of speech output comprises three factors; the users' personal preferences; the operational context; and overall system functionality[17]. These factors are not independent of each other and will integrate to inform the user in their evaluation of speech output.

Personal preferences, operational context & system functionality

Personal preferences for speech are constructed through the application of the social representations held by the listener. Social representations are created by, for example, norms and stereotypes, and are employed to facilitate communication within social groups[18]. Simply based on the 'sound' of speech, social representations are formed in relation to, for example, geographical origin, gender and age.

A large amount of research has focussed on the social representations generated by speech as 'human' output, but few studies consider those generated by speech as system output. The limited number of studies that have investigated social reactions to synthetic speech have discovered that synthetic speech, is often reacted to in similar ways to natural (human) speech. For example, personality and gender are awarded to synthetic speech outputs[19] and gender stereotypes may arise similarly for both natural and synthetic speech[20-22]. Given that speech output is evaluated on the basis of user preferences informed by social representations, it appears appropriate to consider the IDAS users' social representations of speech outputs to determine the spoken characteristics that provoke both positive and negative user evaluations.

The operational context in which speech output is presented also influences the user's evaluation of speech output. User evaluations of IDAS speech output may be based on how the speech output 'fits' with the user's domestic environment. A robotic sounding synthetic speech output may not match the user's perception of their home and they may evaluate a natural speech output more positively. Alternatively, a synthetic speech output may be positively evaluated because of the ease at which it can be discriminated within the environment. However, the influence of the domestic context on users' evaluations of speech output has yet to be explored.

System functionality also impacts user evaluations of speech output. For example, on many lines of the London Underground users are presented with spoken journey information. When asked which voice they would prefer to make such announcements, commuters chose Marilyn Monroe because her voice sounded genuinely helpful[23]. This illustrates how positive user evaluations of speech output were based on matching the characteristics of the speech output (i.e. genuinely helpful) to the system functionality (i.e. to provide help with a journey). Furthermore, it is important that speech outputs provoke consistent

stereotypes with the system's behaviour[19]. Users report that inconsistencies can cause user confusion and distraction leading to negative evaluations of both the system and the speech output. In the case of an IDAS, the system is likely to be perceived as an authority on the alarm states active within the user's home. Therefore a speech output that supports stereotypes of authority could be consistent with the user's perception of IDAS functionality.

Speech source

The majority of studies investigating the evaluation of the sound of speech are restricted to natural speech. However, the limited numbers of studies investigating user evaluation of both natural and synthetic speech outputs have revealed that users may prefer natural speech outputs compared to synthetic alternatives[22]. For example, natural speech is commonly perceived more positively than synthetic speech output[24], with further support from Gong and Lai[25] who indicate that users find synthetic speech unpleasant to listen to.

Other than a strong sense that user evaluation of natural speech output is more positive than for synthetic speech, much of the literature fails to explicitly consider the user's personal preferences, the operational context and system functionality. Therefore it is necessary to investigate user evaluations of speech as IDAS output and to consider these factors of influence.

Speech gender

Even when a speaker is not present, speech is extremely useful in identifying the gender of a speaker and provokes associated social representations[20]. Additionally, Reeves and Nass[19] argue that everyone assigns gender to both natural and synthetic speech outputs. They found that even when the content of the speech output was identical, male speech outputs that praised the user were better received than female speech outputs; male speech output was evaluated as more friendly than female speech outputs; and male speech outputs were evaluated as better information providers on computers. Here, it would be interesting to investigate the current social representations of gender that are invoked in relation to speech as IDAS output and the impact that gender may have on IDAS user evaluations.

Background noise

Given that background noise may effect speech comprehension, it is plausible that it may also impact user evaluation of speech output. The noise generated within an environment may distort the 'sound' of the speech presented and may change the user's evaluation of speech presented in quiet and noisy environments. Within the operational context of an IDAS it is likely that domestic noise will be generated both internally and externally. Although it is documented that spoken background noise has a disruptive effect on cognitive performance[26], how domestic background noise will affect user evaluations of IDAS speech output remains to be explored.

Research hypotheses

Informed by the earlier discussion the following research hypotheses are offered:

(i) *Speech source*: There will be a difference between older adults' evaluations of natural and synthetic speech outputs [$H1_1$];

(ii) *Speech gender*: There will be a difference between older adults' evaluations for male and female speech outputs [$H2_1$];

(iii) *Background noise x speech source*: There will be a difference between older adults' evaluations of speech source

presented in quiet and spoken noise conditions [$H3_1$];
(iv) *Background noise x speech gender*: There will be a difference between older adults' evaluations of speech gender presented in quiet and spoken noise conditions [$H4_1$].

Experimental design

A mixed three-factorial 2x(2x2) experimental design was employed. The between-subjects independent variable was the environmental condition [Quiet, Noise]. The within-subjects independent variables comprised speech source [Synthetic, Natural] and speech gender [Male, Female]. The dependent variable was the participants' evaluations of the speech outputs presented.

Speech evaluations were measured using a response sheet containing the following questions and measurement scales: (i) A question designed to determine whether the participant would like the speech output to be used in their own home, (ii) A semantic differential scale containing six bi-polar attributes; all attributes were directed by previous research[17,19,20,25,27] [Pleasant-Unpleasant, Intelligent-Stupid, Boring-Interesting, Irritating-Soothing, Natural-Unnatural, Clear–Muffled], and (iii) A final question to assess which of the speech outputs was most positively evaluated by the participant.

For both the quiet and noisy environmental conditions identical speech outputs, response sheets and procedure were employed. For brevity, differences between the two conditions are explicitly stated as appropriate.

Participants

32 adults aged 65+, [15 male, 17 female] attending a local day centre took part in the investigation (16 per environmental condition). In the quiet conditions, 7 males and 9 females participated; whereas in the noise conditions, 8 males and 8 females took part. All participants were living independently and presented no severe psychological or physical deficits. All participants spoke English as their first language. Participation was on a voluntary basis and participants were recruited by members of the day centre staff who were familiar with the age and age-related declines presented by their clients. By recruiting in this way screening was avoided, eliminating the need for intrusive enquiries in regards the participants' age and state of physical and psychological health. The investigation took place in a vacant office within the day centre building so as to minimize external (unrestricted) noise from communal areas that could influence the participants' perception of the speech samples.

Speech outputs

An IBM ThinkPad laptop computer was used to present recordings of four speech outputs. The Laureate™ online demonstration (*www.htk.co.uk*), a concatenated speech synthesiser, was used to create both the male and female UK English synthetic speech outputs that were modelled on the Received Pronunciation (RP) accent. The natural male and female speech outputs were recorded using Goldwave™, a digital audio editor and were selected on the basis that both speakers also had RP accents. The same accent was used across speech outputs to eliminate accent as a 'between-voice' discriminatory factor that could obscure the data.

To reflect the IDAS operational context, each speech output presented one of four scripts providing directions around a home environment. Directions were considered a neutral topic that would not invoke participants' preconceived opinions. Each script contained approx-

imately 60 seconds of speech output so that the participant could provide an informed evaluation and would not become fatigued. All speech outputs and scripts were counterbalanced to reduce order effects using the Latin Square technique.

The participants were not asked to perform a concurrent task while listening to the speech outputs. As earlier research within the IDAS domestic context[14,15] had indicated no significant differences between performance data captured for each speech output, it was considered appropriate to present only the speech output type in this instance to ensure that concurrent performance tasks would not influence the participant's subjective evaluation of each speech output type. In this way, we sought to isolate the subjective evaluation variable and consider its impact on the uptake and continued use of an IDAS.

Radio sample (Noise conditions only)
The radio sample used was a recording of '*The Archers*', a UK soap opera broadcast on BBC Radio 4. A radio sample was chosen because of the reported negative effects that competing speech may have on speech perception[26].

Volume levels
Volume levels were dependent on each participant's hearing ability and were considered an ecologically valid approach for speech output delivery. The volume level for all speech outputs remained constant for each participant. When commercially available, IDAS volume levels will be configured in much the same way, due to the diversity and variability of auditory ability within the target user group.

Procedure
At the beginning of each trial the participants were thanked for volunteering to take part in the investigation and ethical considerations concerning their consent, withdrawal and confidentiality were explained. Each participant was given a brief introduction describing an IDAS's general functionality and IDAS-User interaction. It was explained that the current investigation sought to determine which speech output should be used as IDAS output and the speech evaluation task was explained in detail.

The participants were told that they would be listening to four speech output samples; each presenting different directions around a home environment. It was stressed that they should focus on the sound of the speech output, rather than its content. They were informed that after listening to each speech output they would be asked a small number of questions and once all the speech outputs had been listened to, a final question would be posed. They were requested not to compare the speech outputs when answering the questions unless they were specifically asked to do so. Participants in the noise conditions were informed that a radio program would be played during the presentation of the speech outputs.

It was explained that the researcher would ask the questions and complete the response sheets to record the participant's answers. The researcher, taking care to avoid experimenter bias, expanded on the question posed if the participant sought clarity. Once the participant understood the task and was ready to proceed, the task began. It was at this stage in the noise conditions that the radio sample was played. It was stopped only to complete the response sheet after the presentation of each speech output. Playback resumed on presentation of the next speech output. On completion of the task the participants were debriefed and thanked for their co-operation.

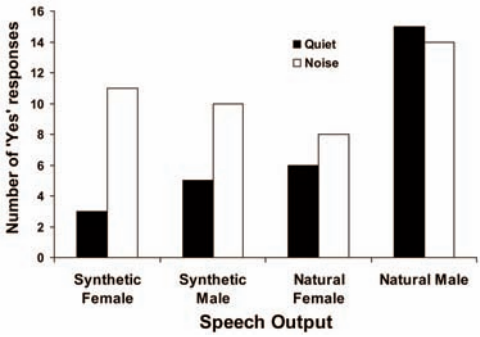

Figure 2. Responses to the question "Would you like this speech output in your own home?"

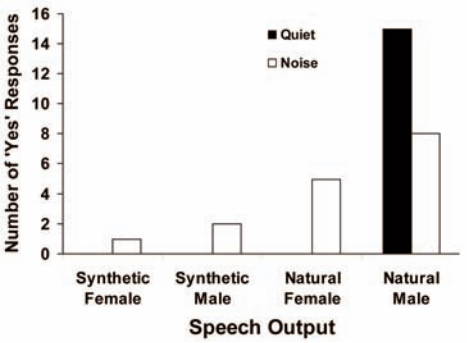

Figure 3. Overall choice for IDAS speech output

Results & analysis

The results of the questions of (i) would you like this speech output to present IDAS output in your own home? and (ii) which speech output overall is preferred as IDAS output? are presented followed by an analysis of the bi-polar attribute data in relation to each research hypothesis.

Speech output suitability for the home?

The affirmative responses to the question of 'would you like this speech output in your own home?' are presented in Figure 2.

Natural speech outputs are favoured over synthetic speech outputs in the quiet conditions (*Figure 2*). However, a mixed picture in terms of preference for speech source is presented in the noise conditions. For speech gender, in the quiet conditions the male speech outputs are favoured over their female counterparts. Again, in the noise conditions these preferences are not as clear.

Overall speech output suitability?

Figure 3 reveals that in the quiet condition 15 participants chose the natural male speech output to be the most appropriate for IDAS speech output. Here, it must be noted that one participant refused to make a decision as he felt that all of the speech outputs were appropriate.

In contrast, the responses provided by the participants in the noise condition revealed that 50% believed that the natural male speech output was most appropriate as IDAS output. 31% chose the natural female speech output, 13% for the synthetic male speech output, followed by 6% for the synthetic female speech output. In the noise conditions the natural speech outputs were chosen as IDAS output by more participants than synthetic speech outputs and male speech outputs were chosen as IDAS output more frequently than the female speech outputs.

Older adults' evaluation of speech outputs

To examine the older adults' evaluations of speech source [natural, synthetic] and speech gender [male, female] in quiet and noisy environmental conditions, an evaluation score for each bi-polar attribute presented within the five-point semantic differential scale was determined. The bi-polar attributes were as follows: *Pleasant*-Unpleasant, *Intelligent*-Stupid, *Boring*-Interesting, *Irritating*-Soothing, *Natural*-Unnatural, *Clear*-Muffled. For means of analysis, evaluation scores were generated for the initial attribute within each bipolar pair (those italicized above) using a scoring system of 1–5 corresponding to where on the five-point scale the parti-

Table 1. ANOVA main effects: Speech source [Natural vs synthetic]

Attribute	dF	F	p
Pleasant	1, 30	6.412	<0.05
Intelligent	1, 30	19.236	<0.05
Boring	1, 30	15.840	<0.05
Irritating	1, 30	6.095	<0.05
Natural	1, 30	49.985	<0.05
Clear	1, 30	11.479	<0.05

Table 3 . ANOVA main effects: Speech gender [Male vs female]

Attribute	dF	F	p
Pleasant	1, 30	5.000	<0.05
Intelligent	1, 30	11.379	<0.05
Boring	1, 30	15.840	<0.05
Fast	1, 30	43.404	<0.05
Irritating	1, 30	14.462	<0.05
Natural	1, 30	7.112	<0.05

Table 2. Post-hoc paired samples t-tests: Speech source [Natural vs synthetic]

Attribute	Synthetic mean	Natural mean	t	dF	p
Pleasant	2.80	2.23	3.066	63	<0.05
Intelligent	2.56	1.97	-4.461	63	<0.05
Boring	2.86	3.53	-4.066	63	<0.05
Irritating	3.06	3.56	2.722	63	<0.05
Natural	2.73	1.84	6.090	63	<0.05

Table 4. Post-hoc paired samples t-tests: Speech gender [Male vs female]

Attribute	Female mean	Male mean	t	dF	p
Pleasant	2.72	2.31	3.066	63	<0.05
Intelligent	2.50	2.03	3.559	63	<0.05
Boring	2.89	3.50	-3.808	63	<0.05
Irritating	2.97	3.66	-3.924	63	<0.05
Natural	2.47	2.11	2.509	63	<0.05

cipant had specified. Thus, a score of 1 can be considered as a very high evaluation of the first attribute within each bi-polar pair and 5 as a very low evaluation. Therefore the lower the evaluation score the higher the evaluation for the initial attribute within each bi-polar pair.

This section reports the 3 factor ANOVA [Environmental condition x (Speech source x Speech gender)] and post-hoc analysis of the bi-polar attribute data in relation to each of the research hypotheses examined.

H1: Older adults' evaluations of natural and synthetic outputs

The mixed 3 factor ANOVA revealed (*Table 1*) that the source of the speech output had a main effect on the evaluations of 'pleasant', 'intelligent', 'boring', 'irritating', 'natural', and 'clear', rejecting the null hypothesis.

Table 2 documents post hoc paired samples *t*-tests revealing that natural speech outputs were evaluated as significantly more 'pleasant', more 'intelligent', less 'boring' , less 'irritating', and more 'natural' than synthetic speech outputs.

A significant interaction effect was observed between speech source and speech gender for the attributes 'clear-muffled'. Post–hoc paired samples *t*-tests revealed that the natural male speech output was evaluated as significantly clearer than the synthetic male speech output [Natural Male M = 1.5, Synthetic Male M = 2.41; t = 4.587, df = 31, p = <0.05] and the natural female speech output [Natural Male M = 1.5, Natural Female M = 2.09; t = 2.600, df = 31, p = <0.05]. Therefore evaluation differences are due to the natural male speech output being evaluated as significantly clearer than the alternative speech outputs presented.

H2: Older adults' evaluations of male and female speech outputs
Table 3 reports that the gender of the speech output had a significant main effect on the evaluations of 'pleasant', 'intelligent', 'boring', 'irritating', and 'natural', thus rejecting the null hypothesis.

Post hoc testing shown in Table 4, using a paired samples *t*-test, revealed that male speech outputs were evaluated as more 'pleasant', more 'intelligent', less 'boring', less 'irritating', and more 'natural' than the female speech outputs.

Older adults' evaluations of speech source [H3ı] and speech gender [H4ı] in quiet and spoken noise conditions
A significant interaction effect was observed between speech source and environmental conditions [quiet, noise] for the bipolar attribute Natural – Unnatural [$F(1,30) = 8.138$; $p = < 0.05$], rejecting the null hypothesis [$H3_0$]. Post-hoc paired samples *t*-tests revealed that synthetic speech outputs were evaluated as more natural in noise conditions than in quiet conditions [Synthetic (quiet) M = 3.41, Synthetic (noise) M = 2.16; $t = 5.683$, $df = 31$, $p = <.05$] and that natural speech outputs were evaluated as more natural in noise conditions than quiet conditions [Natural (quiet) M = 2.06, Natural (noise) M = 1.53; $t = 3.056$, $df = 31$, $p = <.05$]. Therefore the evaluation of both natural and synthetic speech outputs varies when presented in quiet and noise conditions.

Discussion

Overall, the results revealed that the natural male speech output was most positively evaluated as IDAS output by this target user group. Furthermore, the attribute data generally supports this observation with natural and male speech outputs receiving the most positive evaluations compared to their synthetic and female counterparts. The subsequent discussion considers the research hypotheses and explores the impact of speech source [natural, synthetic], speech gender [male, female] and environmental conditions [quiet, noise] on the design of IDAS speech outputs.

Speech Source: Natural vs Synthetic
This investigation revealed that for seven of the eight bi-polar attributes, a significant main effect was observed for speech source indicating that the natural speech outputs were highly evaluated in comparison to the synthetic speech outputs. These results suggest that the natural speech outputs were evaluated as most pleasant, most intelligent, most interesting, most soothing, most natural and most clear, rejecting the null hypothesis ($H1_0$). However, an interaction effect was observed between speech source and environmental conditions for the bi-polar attribute natural-unnatural (rejecting $H3_0$) and will be addressed in the discussion of environmental conditions below. It was also observed that in terms of clarity [clear-muffled] an interaction effect occurred between speech source and speech gender, suggesting that overall differences were due to the high level of clarity awarded to the natural male speech output by the older adults.

In both the quiet and noisy conditions, the natural speech outputs were chosen more often by the participants to be presented in their own home and as most appropriate for IDAS output. These findings support previous research documenting that natural speech outputs are evaluated as more pleasant than synthetic speech outputs[22,24,25]. Simply as a product of their age, older adults are likely to have had relatively more exposure to natural speech than synthetic speech. Thus social representations of natural speech outputs may be deeply rooted and greater exposure to synthetic speech output may be necessary for the opposite effect to occur.

The context in which an IDAS operates may also account for user evaluation of the speech source. The domestic environment is often subject to a variety of natural sources of speech output, such as television and radio. Therefore, natural speech output may be considered more appropriate than synthetic speech output for the domestic context, reflected in positive evaluations of natural speech outputs.

Alternatively, the older adults may have considered the natural speech output more positively based on system functionality. An IDAS is an assistive technology, helping the older adult to remain living independently in their own home for longer. Therefore, the human 'carer' qualities that the system will reflect while supporting the older adult with independent living may have been associated with a natural, rather than a synthetic, speech source.

Speech Gender: Male vs Female

No differences were observed between the older adults' evaluations of male and female speech output in terms of clear-muffled, failing to reject $H2_0$. However, for the remaining five bi-polar attributes a difference was observed between the older adults' evaluations of male and female speech outputs, rejecting $H2_0$. Further analysis of these attributes suggests that older adults evaluated male speech outputs highly in comparison to female speech outputs in terms of pleasant, intelligent, interesting, soothing and natural. The participants' responses to the additional questions posed support the evaluations of the male speech outputs reported. Male speech outputs were consistently judged more appropriate than female speech outputs in both noisy and quiet conditions when the participants were asked 'would you like this speech output in your own home?'

These findings appear to support Reeves and Nass[19], and their position that male speech output is more positively evaluated than female speech output when presenting technical information. It is likely that participants perceived an IDAS as a technical system. Therefore it is possible that the delivery of technical information may have influenced the overwhelmingly high evaluations of the male speech outputs. Furthermore, given the safety-critical functionality of an IDAS, the participants may have felt that the male speech outputs connoted greater levels of 'authority'. There is, of course, the possibility that the male speech outputs were easier to understand. However, earlier research, reported elsewhere[14], suggests that this was not the case.

Environmental Conditions: Quiet vs Noise

The environmental conditions were observed to have had an effect on the older adults' evaluations of the speech outputs in terms of boring-interesting and clear-muffled. The older adults evaluated the speech outputs as highly boring in the quiet conditions compared to the noise conditions. In terms of clarity, the older adults evaluated the speech outputs as possessing higher levels of clarity in the noise conditions than in the quiet conditions. These findings are surprising as it was expected that rather than induce positive evaluations, the presence of background noise would negatively affect the speech output evaluations.

Interaction effects between environmental conditions and speech source were observed for the attributes natural – unnatural, rejecting the null hypothesis $H3_0$. The interaction effect revealed that both natural and synthetic speech outputs were evaluated as more natural in the noise conditions than in the quiet

conditions. In contrast, no interaction effects were observed between environmental conditions and speech gender for any of the attributes. This suggests that the environmental conditions did not have an effect on the older adults' evaluations of male and female synthetic speech outputs, failing to reject the null hypothesis ($H4_0$).

The observed interactions between environmental conditions and speech source in terms of 'naturalness', may be explained by the 'naturalness' of speech relying heavily on the tonal qualities of the speech output. The variation between speech output evaluations in the quiet conditions may not have been observed in the noise conditions due to the background noise producing a distortion/masking of the speech source, so that the older adults observed little difference in tonal quality.

Design recommendations & methodological reflections

This investigation has revealed that overall, natural and male speech outputs are most positively evaluated by older adults when considered as IDAS output. Therefore it is recommended that, by default, a natural male speech output should be used as IDAS output. Furthermore, given that evaluation differences were observed between speech sources and speech genders, future system designers must consider subjective evaluations of speech output with equal weight of importance to that of speech comprehension if system uptake and continued use is to be promoted.

Previous research suggests that a structured, systematic approach to investigating design requirements for older adults should be used[28]. This recommendation was based on observations of focus groups comprising older adults and their tendency to exhibit behaviours, such as frequent diversions from the topic under discussion, that did not lend themselves well to the elicitation of appropriate data that could be used to inform system designs. In this earlier work, it was recommended that methods that employed control, yet allowed for some flexibility, such as semi-structured interviews, should be used where in-depth qualitative data is required. However, in this instance an experimental methodology was chosen based on the demands of the inquiry (i.e. the measurement of several independent variables) and the fixed structure and systematic processes that an experimental design offers. On reflection, the older adults participating in this investigation appeared to respond well to the 'rules' of the experiment and were keen to follow direction. In a few cases, during pauses between questions, the same tendencies to actively instigate conversation that was unrelated to the current task were observed. Experienced in conducting research with older adults, the researchers managed such distractions well by stating that they could have a 'chat' at the end of the experiment (with all participants in this instance being willing to do so). Initially, it was considered that the conversational interruptions offered by the older adults were due to fatigue and that the experimental design should be adapted to allow for 'rest' periods. However, participants assured the researcher that this was not the case and that they were enjoying the experience.

In light of this, it is recommended here that 'chat' time should become a formal experimental design feature, and is formally highlighted during the experimenter's introduction to the task and the procedure. By incorporating this feature, the ageing user group may be less inclined to strike up conversation at inappropriate times that could lead to

data inaccuracies/distortions, enjoy the 'experimental' experience and have their conversational/social inclusion needs met. Thus providing an equally productive and enjoyable experience for both the researcher and participant.

Future Research Directions

Future research will seek to address key limitations of the current investigation. Field experiments within the domestic context will be conducted to determine whether the subjective evaluations made in the laboratory are reflected within the IDAS operational context and whether such evaluations remain the same over time.

Acknowledgements

This research was supported by the Millennium Home Project (EPSRC/DTI Link Foresight/DoH Medlink). We would also like to thank Hillingdon Social Services for their cooperation.

References

1. Dunnell K. Policy responses to population ageing and population decline. United Kingdom. Presented at The Expert Group Meeting on Policy Responses to Population Ageing and Population Decline. Population Division, Department of Economic and Social Affairs, United Nations Secretariat, New York, USA; 2000
2. Help The Aged. Nothing Personal: Rationing Social Care for Older People; 2002, *www.helptheaged.org.uk/NR/rdonlyres/epynf7emqakgkwyyacbszvo63p7b32ylvedgva7qvmnb6lnsdkbngofbsvvdzn2t66k44qbjuantvexl2hpbhdnutvg/Nothing_Personal.pdf*, accessed 12-01-2006
3. Berlo A van. (2001). Design Guidelines on Smart Homes. A COST 219bis Guidebook. Finland: STAKES; 2001; *www.stakes.fi/cost219/smarthousing.htm*, accessed January 12, 2006
4. UK Department of Health. Research and Development Work Relating to Assistive Technology 2004-05. London: HMSO; 2005, *www.dh.gov.uk/assetRoot/04/08/65/92/04086592.pdf*, accessed January 12, 2006
5. Monk A, Hone K, Lines L, Dowdall A, Baxter G, Blythe, M, Wright, P. Towards a practical framework for managing the risks of selecting technology to support independent living. Applied Ergonomics; in press
6. Gregor P, Newell AF, Zajicek M. Designing for Dynamic Diversity – interfaces for older people. In: Proceedings of ASSETS 2002, The 5th International ACM Conference on Assistive Technologies, July 8-10, Edinburgh, UK; 2002; pp 151-155
7. Hawthorn D. Possible implications of aging for interface designers. Interacting with Computers 2000;12:507-528
8. Jacko JA, Scott IU, Sainfort F, Barnard L, Edwards PJ, Emery VK, Kongnakorn T, Moloney KP, Zorich BS. Older adults and visual impairment: What do exposure times and accuracy tell us about performance gains associated with multimodal feedback? In: Proceedings of CHI 2003, April 5-10, Florida, USA; 2003;5(1):33-40
9. Emery VK, Edwards PJ, Jacko JA, Maloney KP, Barnard L, Kongnakorn T, Sainfort F, Scott IU. Toward achieving universal usability for older adults through multimodal feedback. In: Proceedings of ACM Conference on Universal Usability, 10-11 November, Vancouver; 2003; pp 46-53
10. Perry MJ, Dowdall A, Lines L, Hone K. Multimodal and ubiquitous computing: Supporting independent-living older users. IEEE Transactions on Information Technology in Biomedicine 2004; 8(3):258-270
11. RNIB. Office of National Statistics mid-1996 population estimates. London: RNIB; 2000, *www.rnib.org.uk/wesupply/fctsheet/authuk.htm*, accessed January 12, 2006
12. Hone KS, Graham R, Maguire MC, Baber C, Johnson GI. Speech technology for automatic teller machines: an investigation of user attitude and performance. Ergonomics 1998;41(7):962-981
13. Gibbon D, Mertins I, Moore RK. Handbook of Multimodal and Spoken Dialogue Systems, Resources, Terminology and Product Evaluation. Norwell: Kluwer; 2000
14. Lines L, Hone KS. Older Adults' Comprehension of Speech as Interactive Domestic

Alarm System Output. In: Proceedings of ECCE 11, 11th European Conference on Cognitive Ergonomics, September 8-11 2002, Catania, Sicily; 2002; pp 335-340
15. Lines L, Hone KS. Interactive Domestic Alarm Handling: Older Adults' Interaction with Speech Alarms. In: Proceedings of the 26th International Conference on Information Technology Interfaces ITI 2004, June 7-10, Cavtat, Croatia; 2004; pp 367-372
16. Sanders MS, McCormick EJ. Human Factors in Engineering and Design. 7th Edition. Singapore: McGraw-Hill; 1993
17. Salame P, Baddeley AD. Disruption of short-term memory by unattended speech: Implications for the structure of working memory. Journal of Verbal Learning and Verbal Behaviour 1982;21:150-164
18. Moscovici S. Social Influence and Social Change. London: Academic Press; 1976
19. Reeves B, Nass C. The Media Equation. How People Treat Computers, Television, and New Media Like Real People and Places. New York: Cambridge University Press; 1996
20. Mullennix JW, Stern SE, Wilson SJ, Dyson C. Social perception of male and female computer synthesized speech. Computers in Human Behaviour 2003;19:407-424
21. Nass C, Robles E, Heenan C, Bienstock H, Treinen M. Speech-Based Disclosure Systems: Effects of modality, gender of prompt and gender of user. International Journal of Speech Technology 2003;6:113-121
22. Stevens C, Lees N, Vonwiller J, Burnham D. On-line experimental methods to evaluate text-to-speech (TTS) synthesis: effects of voice gender, and signal quality on intelligibility, naturalness and preference. Computer Speech and Language 2005;19:129-146
23. Anonymous. In brief: Marilyn to go underground? The Guardian Newspaper, Home News; 1999(May 7):16
24. Stern SE, Mullennix JW, Dyson C, Wilson SJ. The persuasiveness of synthetic speech versus human speech. Human Factors 1999;41:588-595
25. Gong L, Lai J. To mix or not to mix synthetic speech and human speech? Contrasting impact on judge-rated task performance versus self-rated performance and attitudinal response. International Journal of Speech Technology 2003;6:123-131
26. Salame P, Baddeley AD. Disruption of short-term memory by unattended speech: Implications for the structure of working memory. Journal of Verbal Learning and Verbal Behaviour 1982;21:150-164
27. Simpson C A, McCauley ME, Roland EF, Ruth JC, Williges BH. Speech controls and displays. In: Salvendy G, editor, Handbook of Human Factors. New York: Wiley; 1987; pp 1490-1525
28. Lines L, Hone KS. Eliciting user requirements with older adults: Lessons from the design of an interactive domestic alarm system. Universal Access in the Information Society 2004;3(2):141-148