The Effect of Age and Native Speaker Status on Intelligibility

Catherine Watson[1], *Wei Liu*[1], *Bruce MacDonald*[1]

[1] Department of Electrical and Computer Engineering, University of Auckland, New Zealand

c.watson@auckland.ac.nz,b.macdonald@auckland.ac.nz

Abstract

We investigate whether listener age or native speaker status has the biggest impact on the intelligibility of a synthetic New Zealand English voice. The paper presents findings from a speech intelligibility experiment based on a reminding task involving 67 participants. There were no significant differences in the results due to age (young and old adults), however there was for native speaker status. The non-native listeners performed significantly worse than the native listeners in the synthetic speech condition although no differences were found in the natural speech condition. We argue that despite the fact that aging impacts on speech perception, the older native listeners were able to draw on their in depth language model to help them parse the synthetic speech. The non-native speakers do not have such an in depth model to assist them.

Index Terms: speech synthesis intelligibility, older listeners, non-native listeners.

1. Introduction

Speech is increasingly used as an interface between humans and machines, particularly by social and social assistive robots ([4-14]). One such example is the HealthBots project which has developed social assistive robots for older adults ([10-13]). These robots have been deployed extensively in a New Zealand retirement village. They provide a broad range of services to support the daily activities of the users, including vital signs measurement (e.g. pulse rate, blood pressure and glucose level), schedule reminding (medication, and meetings), telepresence communication, human-robot interaction (e.g. greeting, authentication), and entertainment in the form of jokes and music. The robot communicates to users via a computer screen and a synthetic voice, and the users communicate via buttons on a touch screen.

There is a significant body of work on human-robot interfaces, particularly targeted at the aged population (see for example [7-14]). Acceptance of these healthcare robots by users is crucial for the successful deployment of robotics for aged care. In the HealthBots project, three extensive trials on the feasibility and quality of life effects of the robots in the retirement village have been performed over the last three years (see for example [11-13]). It was found that participants rated the robot highly in terms of overall quality of Human Robot interfaces [12], suggesting participants found the robot acceptable. There were also improvements in attitude towards the robots after meeting the robot [12] and the robot voice plays an important part in its acceptability [17].

1.1. The New Zealand English Voice

A diphone-based male New Zealand English (NZE) voice, running with the Festival Speech Synthesis framework [15] has been developed for the robots [16-19]. An NZE pronunciation dictionary was developed in conjunction with this voice, which contains not only words specific to the NZE lexicon (such as many Maori loan words), but also NZE specific pronunciations (for example the merger of all /iə/ and /eə/ diphthones to /eə/) [16]. The choice of using a New Zealand accent was a deliberate one. People infer much about a person from their accent (e.g. likeability, authority, trustworthiness)[21-23], and negative perception of a voice can cause confusion in communication intent [23]. In socially interactive technologies there is also evidence that the accent influences the impression and acceptance of the interface [eg 17,23,24]. In a New Zealand based study we found people reacted significantly more positively to the robot if it had our NZE voice as opposed to the US accented KAL voice from Festival [17] and in addition they thought the performance of the robot was better. This was despite the fact the task and dialogue the robot performed was exactly the same, only the voices differed. In [18] we specifically tested the perception of speech quality of the NZE voice by older people in a New Zealand retirement village. We found the NZE voice was ranked higher than the KAL voice, and participants reported that they felt they could listen to it longer.

Through speech quality tests we have received feedback on how pleasant people found the NZE robot voice [16-19] and anecdotally via comments collected from the robot trials. Age does effect the perception of quality, with the older listeners giving the voice lower speech quality scores than the younger listeners [18]. These speech quality tests have helped inform us on what improvements are needed for the voice, particularly for the older users. For instance we created a new Festival function that enables different vocal affect, in response to the request to make the voice less robotic [19]. However, there has not yet been a rigorous test of the intelligibility of the NZE voice. The speech is clearly intelligible as the robots have been successfully employed in the trials. But it is important to quantify the intelligibility to identify what areas of the synthetic voice need improvements.

There are user specific issues associated with speech intelligibility for the HealthBots robot. Speech intelligibility, whether it be synthetic or natural, is dependent on factors such as hearing status, listeners age, native speaker status (e.g [25-37]). These latter two are of specific importance to the HealthBots project due to the language background of the two the main user groups of our healthcare robot. These two groups are New Zealand senior citizens (who are currently predominantly NZE speakers), and their carers, many of whom are second language speakers of English [38]. There may be speech intelligibility issues associated with both groups, and we need to have a better understanding about these.

. This paper outlines our investigation into the impact that both age and native speaker status have on speech intelligibility for both the synthetic NZE voice, and a natural voice. In section 1.2 we give a brief overview of other intelligibility studies with relevant findings to our study here. In Section 2 we present the findings of our own study, and discuss the implications of them in Section 3, before concluding in Section 4.

1.2. Speech intelligibility

There is strong evidence that the aging process impacts on speech perception, whether it be the ability to resolve differences in segmental temporal cues (e.g.[25]), understanding speech in background noise (eg. [26]), or, due to degradation in process resourcing (eg. [33]). These difficulties are expected to be further compounded in synthetic speech [27-30]. Studies consistently show that natural speech is more intelligible than synthetic speech [eg. 29-32]. Further, older participants perform consistently worse in intelligibility tests with synthetic speech, compared to younger adults [29-33]. In particular Wolters and colleagues [eg 29,30] demonstrated that older participants make more errors compared to younger participants in recalling complex words, such as medicine names.

However, regardless of age, Native speakers do have a vast knowledge of the sound system of their language, and for degraded speech they are able to exploit the fact there is a large amount of redundancy in spoken language, which is in contrast to the non-native speaker [28,36,37]. Being a non-native speaker is known to impact on intelligibility of synthetic speech [27,32,34], although the level of skill in the host language can reduce the impact [27]. However it is of note that Jones et al [28] found no difference in the comprehension rates between native and non-native speakers of English for a synthetic Australian English voice.

To date there has been no study that contrasts the synthetic speech intelligibility scores by listeners who are older adults to those who are non-native. In [32] they found older listeners performed worse than younger listeners in a synthetic speech intelligibility task. When they then split their young listeners group into native and non-native speakers, they found the non-native speakers performed worst. Perhaps the older first language listeners would have performed worse than younger non-native listeners, but this was not tested.

In this study we will explicitly look at synthetic speech intelligibility issues contrasting older and younger listeners, and native and non-native listeners. We expect greater accuracy with stimuli from natural speech than synthetic speech, and from less complex stimuli than complex stimuli. We also expect that the older listeners would find the intelligibility task harder than the younger listeners , and the non-native speaker listeners to find the task harder than native speaker listeners.

2. Speech Intelligibility Study.

We based our Speech Intelligibility study on a schedule reminding task proposed by Wolters et al ([29,30]). This tasked involved reminders about meeting people, and taking medicines. This experiment was ideally suited to our study, as reminding is one of the main services of the HealthBot Robots. In the exercise, participants were given a reminder about either meeting a person or taking medication at a specified time. They then were then required to identify the time the action was to take place, the person's name, or the medication name. For this study the robots were not involved. Participants only heard the voice via an online survey. We did try to recruit participants from the retirement village, where the robots were being tested, however this proved very difficult. This survey came immediately after the three robot trials and the residents most likely had experiment fatigue. By having the survey online, we were able to recruit participants beyond the retirement village.

2.1. Methodology

In the study participants were given reminders in either a natural NZ voice or the synthetic NZE voice and we compared participant's abilities to correctly recall the reminders. Approval was granted for this study by The University of Auckland Human Participation Ethics Committee.

2.1.1. The stimuli

The reminders the participants were given are listed in a generalized form in Table 1. Table 2 lists the times, people names, and medicine names used in the recall task. For the times we state the time of the day (morning, afternoon, evening), rather than stating AM and PM as in [29,30] as we believe this is more natural in spoken speech. In addition the names of the people and medicines were also adapted from [29,30] to be more applicable to New Zealand.

Table 1 The general form of the reminder stimuli. These were first proposed in [29,30].

Reminder	Template
Meeting Reminder	At "time", you are meeting PERSON. You are meeting "person" at TIME.
Medication Reminder	At "time", you need to take your MEDICATION. You need to take your MEDICATION at "time".

Table 2 The times, names of people, and names of medicines used in this study, adapted from [29,30].

Category	Stimuli
TIME	5:05 in the afternoon, 2:25 in the afternoon, 8:45 in the afternoon, 6:40 in the evening, 11:40 in the morning, 8:20 in the morning, 8:10 in the evening, and 2:35 in the afternoon.
PERSON	Judy, Julie, Nicky, Ricky, Kim, Jim, Ted, and Ned
MEDICATION	Accumycin, Omeprozole, Beclotor, Cilazapril, Colecalciferol, Dexozine, Digoxin and Felodopine.

The focus items for this study are capitalized in Table 2. Participants would hear a reminder, and then were required to respond to a questions about the reminder (e.g. for the reminder "You are meeting Jim at 5:05 in the afternoon", they would be asked: "What time will you meet Jim?". For the synthetic NZE voice the focus word was emphasized by placing a small pause immediately before to it, along with a slight pitch rise syllable prior to the focus word.

For both voices the questions were formulated so the focus item was usually in the second part of the sentence, but for medication names we also asked participants to recall these when they fell in the first part of the sentence. In a preliminary study [35] we found, as in [29,30], that participants recalled items better if they occurred in the second part of the sentence in the reminder. There was no need, therefore, to retest this thoroughly in the larger study.

There were 32 reminders in total: 8 TIME and NAME reminders, and 16 MEDICINE reminders. The order of the reminders was randomized. Two versions of these 32 reminders were compiled; one with natural speech (also an NZE male speaker), and one with the NZE voice. The text content was exactly the same.

2.1.2. *The web-based survey*

The entire intelligibility test was delivered online using the open source software LimeSurvey (www.limesurvey.com) on a web server at The University of Auckland. Participants were required to have a playback facility on their computer. At the beginning of the survey, the web service provided participants with a client-system test to ensure that both the survey website and the multimedia player were running correctly on the user's PC. Prior to the Speech intelligibility task, participants were given an example question and answer. Participants typed their response to each reminder into a text box. They were only able to listen to each reminder once. The intelligibility part of the survey took about 20 minutes to complete. There was a speech quality survey as well [35]. This took about 10 minutes complete and was presented first. Only the results of the intelligibility test are presented here.

2.1.3. *Participants*

Eighty-one participants were recruited for the experiment. Unfortunately 12 participants only partially completed it. Most likely this was due to either participant fatigue or issues with the internet connection, judging by their comments. Of the remaining 69 participants 50 were in the young group, aged between 18-25 years, and 19 were in the older group, aged over 45 years. The majority the older group was aged over 60 (12/19). In addition 43/69 participants had English as their first language (henceforth referred to as the L1 group), and 26 were second language speakers of English (henceforth referred to as the L2 group). Twenty-four of the L2 group were from the young group. They had lived in an English-speaking country for a mean of 9.8 years, with a standard deviation of 4.8 years. The remaining two L2 participants were in the older group. However we removed these two from the experiment, as this wasn't a sufficiently high enough number of participants for the old L2 category to be experimentally viable. The very low number in the old L2 group was due to the fact we did not specifically target this group, but the online survey was open to everyone.

Given the task was online; it was not feasible to do hearing tests, however participants did have to fill out a questionnaire about themselves, which included a question on hearing. Four of the older group reported hearing problems, none of the younger group did. The participants heard the reminders being given in one voice type only, and were randomly allocated to one of the two categories (natural or synthetic). Table 3 below lists the age, language background and number of participants hearing each voice.

Table 3 the number of participants who heard each voice type, according to age and language experience.

Voice Used In Task	Young		Old
	L1	L2	L1
Natural	10	9	4
NZE voice (NZE)	16	15	12

2.1.4. *Analysis of Data*

All responses were scrutinised and compared to the intended word. It was necessary to use a certain amount of discretion to decide whether a response was correct or not. Participants were only asked to write down what they heard and were assured that it was not necessary to spell the word exactly right. There was very little variation in the spelling of people's names, but there was for the spelling of the medicine, and any reasonable phonetic spelling was accepted (for e.g. for the Drug Accumycin acceptable answers included "Ackumisin", "Acumysene", "Accumycin", unacceptable answers included "Attemeisin" "Aclimentin: "accumice"). All times where entered in a digit format and an indication of the time of day (e.g. 5:05 pm). All responses were marked as correct or incorrect, and then the package R [39] was used to do the statistical analysis.

2.2. Results

A multi-way ANOVA was used to analyse the data. The number of correct responses was the dependent variable, and there were three independent variables: participant type (older L1 adult , younger L1 adult, younger L2 adult), voice (natural vs. synthetic), and Stimuli type (TIME, PERSON's name, MEDICINE). All three independent variables were found to be significant. Posthoc Tukey HSD tests, and t-tests were then used to further examine the results. These are discussed in detail below.

2.2.1. *The effect of voice*

There were more correct responses for the natural voice (mean score 76.8%, SD 20.7%) than for the synthetic NZE voice (mean score 71.6 %, SD 24.3%), and this was significant ($F(1,261)=4.5, p>0.05$).

2.2.2. *The effect of L1 vs L2*

There were significant differences in the responses of the three different participant categories ($F(2,261)=11.9$, $p<1e\text{-}4$). The difference was an L1 vs L2 effect only. There was no significant difference between the two L1 groups (Posthoc Tukey p=0.9), the mean correct score for the old L1 was 77.9% (SD 25.1%), and for the young L1 was 77.1% (SD 19%). In contrast the L2 Young group (mean correct score of 66.0 %, SD=24.1%) performed significantly worse than the L1 Young group (Posthoc Tukey p <1e-3), and the L1 Old group (Posthoc Tukey p<1e-3). For age related effects it is only appropriate to compare the two difference L1 groups, and for this study there is no significant difference. Although it is interesting that the L1 old group was more variable than the young L1 group.

Table 4 L1 and L2 groups Mean Correct scores for all the stimuli (*left*),NZE voice(*middle*),Natural (*end*).

	Total	NZE	Natural
L1 group	77.5%	76.6%	79.2%
L2 group	66.0%	61.0%	72.9%

The mean correct scores for the L1 and L2 participants for all the stimuli, both Natural and Synthetic, are given in Table 4. The results of the young and old L1 groups have been pooled together since there was no significant difference between the two. The mean correct scores from the Natural speech are much higher than for the synthetic speech, which is expected since the Voice factor was significant. Interestingly although the L2 group performed worse than the L1 group for both the Natural and Synthetic conditions, only the differences in the Synthetic condition are significant. Posthoc t-test (Bonferoni corrected for multiple comparisons) showed that the L2 responses for the synthetic stimuli were less than both those from the young L1 group ($t[112.5]=3.8, p<1e-2$), and the old L1 ($t[104.9]=3.0, p<0.05$). In addition the t-tests showed that there were no differences between any of the three groups in the natural speech conditions.

2.2.3. *The effect of Stimuli*

Table 5 Results of the intelligibility of different stimuli in a sentence

Stimuli	Overall	L1	L2
FM	59.7%	65.1%	50.0%
SM	62.9%	66.0%	57.3%
Name	87.1%	91.9%	78.6%
Time	83.8%	86.9%	78.1%

The stimuli type also impacted significantly on intelligibility, as can be seen in Table 5 ($F(3,261)=36.8, p=0$), where FM stands for the medication name in the first, or earlier, position in the sentence, and SM stands for the MEDICATION name in the second, or latter, position. All of the NAME and TIME stimuli were located in the second position in the sentence. Posthoc Tukey HSD tests on the overall results showed that both the NAME and TIME stimuli have a significant higher intelligibility than medication in either position in a sentence (NAME–FM: diff = 27.4%, p = 0; SM–NAME: diff = -24.3, p = 0; TIME–FM: diff = 24.1%, p = 0; TIME–SM: diff = 20.9%, p = 0). There is no notable difference between the intelligibility of MEDICATION stimuli located in either position in the sentence (SM–FM: diff = 3.2, p = 0.77), nor any intelligibility differences between TIME and NAME stimuli (TIME–NAME: diff = -3.4, p = 0.74).

From our earlier findings (see section 2.2.2) we would expect the L2 participants to do worse than the L1 participants. This can be seen clearly in Table 5, across all 4 stimuli groups. However there was no significant interactions between the stimuli type and language background of the participants, which means the tasks the L1 listeners found easy were also found to be easy by the L2 listeners (i.e. identifying the names and times), and the tasks the L1 listeners found hard, were also found hard by the L2 listeners (i.e. identifying medicine names).

3. Discussion

Some of the findings of this study reinforced our earlier expectations, however some were unexpected. As expected the participants recalled items better if they heard the reminders in the natural voice, as opposed to the synthetic voice. This was also found in [29-31,33]. We note though, that there was only an absolute difference of 5% in the overall intelligibility between the natural and synthetic speech. We also found that less complex stimuli (in our case times and peoples' names) were much easier to recall than the complex ones (in our case medicine names), this was also found in [29-31]. However in contrast to [29,30], we found there was no significant difference in the ability to recall a medicine name regardless of its position in a sentence. It is possible that our method of increasing focus on the keyword (see Section 2.1.1) may have aided in reducing the impact of sentence position on the keyword.

The most significant results from this study are that native speaker status has a greater impact on the speech intelligibility than age. In fact the old L1 speakers performed at the same level as the young L1 speakers, this also is unexpected.

One possible explanation for the high performance of the old L1 group was they were not old enough. The oldest L1 group in this study comprised 4 in the over 75 years group and 8 in the 60-75 years group, however there were also 7 between 45-60 years (although most in this group were nearly 60 years). However when we redid the ANOVA, splitting the L1 group into three groups (the young L1 speakers (as before), those between 45-60 years (7 in this group), and those 60 years and over (12 in this group)) the same variables remained significant. Further, there remained no significant difference between the young L1 speakers (77.1%) and old L1 group - now aged over 60 years old (mean 73.4%).

Other studies have indicated that it is the hearing status of the participants, not the chronological age that is the most important influence on performance [eg. 29,30,33]. The over 60 years L1 group, had all the self–reported cases of hearing impairment, and yet their intelligibility scores were still not significantly different from the young L1, although their mean score is less than the young L1 speakers. Perhaps by increasing the participant numbers this difference would become significant. However the relatively good performance of the older group is further evidence that our methods employed to improve the intelligibility of the voice are working. As is the fact that no significant differences was found for the L1 speakers between their speech intelligibility scores for the synthetic NZE voice and the natural voice (see section 2.2.2).

It is also very noteworthy that there was no significant difference between the L1 and L2 group in the natural voice condition (with a mean of 79.2% and 72.9% respectively, see Table 4). The L2 group only perform significantly worse than the L1 group in the synthetic speech condition (mean 61.9% vs 76.6%, see Table 4). This suggests that it is not the complexity of the reminding task, in particular the spelling of the medicine names that is causing the comparative low score for the L2 group, it is the processing the synthetic speech.

Previous speech intelligibility studies have attributed the poorer performance of the L2 participants, compared to the L1 participants, to less depth of knowledge in both the sound system and linguistic structure of their second language [27, 28, 36]. By contrast the poor performance of older participants compared to young participants in intelligibility studies has been attributed to hearing issues [eg. 29,30,33] and memory issues [eg. 29,30]. However, as noted by [28,36, 37] adult L1 listeners have an in depth knowledge of the sound system and linguistic structure of their language. This helps with intelligibility in natural speech; in [36] they showed L1 listeners benefit from clear speech, but L2 listeners do not. Our study here suggests that whilst the synthetic speech increases the cognitive load on the listeners, the L1 listeners are able to draw on their language knowledge in a way the L2 cannot. Thus whilst the older L1 participants would have all had age related hearing loss [26] (some to the level it was self-reported) their in depth language knowledge helps compensate for this. A natural conclusion from this study would be that a group of older L2 participants would perform the worst at this task. As we only had two participants in this category, we are not yet in a position to test this, once again more participants would be required.

However for the HealthBots project the findings from this experiment are significant. The numbers of overseas born careers for the New Zealand Elderly have been steadily increasing, between 1991 and 2006 the numbers of careers born overseas increased from 19% to 25% [38], and this increase continues [40]. But more significantly the number of careers for whom English was not their first language has also increased dramatically [40].Therefore the fact that the young L2 group performed significantly worse than either of the L1 groups in our study has significant implications for human-machine interfaces in the HealthBots project, and in fact for all human machine interfaces in healthcare of the Elderly. This is because throughout the western world there is an increasing reliance on carers whose first language is not the language of the country in which they live [38].

In the HealthBots project the focus to date has been improving the intelligibility of our Healthcare robot voice for the Elderly users. We knew that speech intelligibility of the robot would also be an issue for their carers (many of whom have English as a second language), however we made the assumption that age would be a bigger disadvantage than language background. The result of this study suggest this assumption, has been wrong. Therefore although issues with the elderly population remain our main focus, it is now clear we need to also perform our intelligibility tests on L2 speakers too.

4. Conclusions

In this study we have shown that in a reminder task complex words are harder to recall than simple words, reinforcing the findings of [29-31]. The intelligibility of the NZE voice compares favourably to natural speech; in fact for the First Language listeners regardless of age, there was no significant difference between the intelligibility of the two voice types. However for the Second Language listeners the NZE voice is significantly less intelligible than the natural voice. Consequently the main findings of this study, is that for the synthetic NZE voice it is the native speaker status of the listener, not age that has the biggest impact on speech intelligibility. This has major implications for human-robot interfaces in healthcare owing to large number of increasing number of carers in the health systems who are second language speakers in their country of employment.

5. Acknowledgements

We thank the participants of the study, the HealthBots project members, and the New Zealand Ministry for Science and Innovation for funding the HealthBots project. We also thank ETRI for their contributions to the HealthBots project.

6. References

[1] Murray, M. K, "The Nursing Shortage: Past, Present, and Future,"Journal of Nursing Administration, vol. 32, p. 79, 2002.

[2] Super, N., "Who will be there to care? The growing gap between caregiver supply and demand," in National Health Policy Forum, George Washington University, Washington DC, 2002.

[3] United Nations, World Population Prospects: The 2006 revision. New York: United Nations, 2006.

[4] Ichbiah, D., Robots: From science fiction to technological revolution. New York: Harry N Abrams, 2005.

[5] Krebs, H.I., Palazzolo, J.J Dipietro, L.,Ferraro, M., Krol, J., Rannekleiv, K.,Volpe, B.T and Hogan, N. "Rehabilitation robotics: Performance-based progressive robot-assisted therapy," Autonomous Robots, vol. 15, pp. 7-20, 2003.

[6] Granata, C., Chetouani, M., Tapus, A., Bidaud P.and Dupourqué, V., "Voice and graphical-based interfaces for interaction with a robot dedicated to elderly and people with cognitive disorders", in Proc. of the 19th IEEE International Symposium on Robot and Human Interactive Communication, pp.785–790, 2010.

[7] Mataric, M.J., Eriksson, J., Feil-Seifer, D.J. and Winstein, C.J., "Socially assistivee robotics for post-stroke rehabilitation," J. of NeuroEngineering and Rehabilitation, vol. 4, p. 5, 2007.

[8] Ellison, L.M, Pinto, P.A, Kim, F., Ong, A.M, Patriciu, A., Stoianovici, D., Rubin, H., Jarrett, T., and Kavoussi, L.R., "Telerounding and patient satisfaction after surgery," Journal of the American College of Surgeons, vol. 199, pp. 523-530, 2004.

[9] Pollack, M., Engberg, S., Matthews, J.T., Thrun, S., Brown, L., Colbry, D., Orosz, C., Peinter, B., Ramakrishnan, S.,Dunbar-Jacob, J., McCarthy, C., Montemerlo, M., Pineau, J., and Roy, N., "Pearl: A Mobile Robotic Assistant for the Elderly," Workshop on Automation as Caregiver: the Role of Intelligent Technology in Elder Care, (AAAI), August 2002.

[10] MacDonald, B., Abdulla, W., Broadbent, E., Connolly, M., Day, K., Kerse, N., Neve, M., Warren, J., and Watson, C.I., "Robot assistant for care of older people," in Proceedings from the 5th International Conference on Ubiquitous Robots and Ambient Intelligence, 20–22 November 2008.

[11] Jayawardena, C., Kuo, I.H., Unger, U., Igic, A., Wong, R., Watson, C.I., Stafford, R.Q., Broadbent, E., Tiwari, P., Warren, J., Sohn, J., MacDonald, B.A., "Deployment of a service robot to help older people," in Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots & Systems (IROS), pp.5990-5995, Oct. 2010.

[12] Stafford, R.Q, Broadbent, E., Jayawardena, C.,Unger, U., Kuo, I.H., Igic, A., Wong, R., Kerse, N., Watson, C.I., MacDonald, B.A, "Improved robot attitudes and emotions at a retirement home after meeting a robot,"2010 IEEE RO-MAN, pp.82 – 87, Sept. 2010.

[13] Jayawardena, C., Kuo, I., Datta, C., Stafford, R.Q., Broadbent, E., and Macdonald, B.A., "Design, implementation and field tests of a socially assistive robot for the elderly: HealthBot Version2" 4th IEEE RAS/EMBS Int. Conf. on Biomedical Robotics and Biomechatronics Roma, Italy. Pp1837-1842, 2012.

[14] Giuliani, M.V., Scopelliti, M and Fornara, F. "Elderly people at home: technological help in everyday activities," in IEEE international workshop on robots and human interactive communication, USA, 2005, pp. 365-370.

[15] Black, A., Taylor, P., and Caley, R., "The Festival speech synthesis system," 1998. Online: http://www.cstr.ed.ac.uk/projects/festival.html.

[16] Watson, C.I, Teutenberg, J., Thompson, L., Roehling, S., and Igic, A., "How to build a New Zealand voice," in NZ Linguistic Society Conference, Palmerston North, Nov30 – Dec 1 2009

[17] Tamagawa, R. Watson, C.I., Kuo, I.H., Macdonald, B.A., and Broadbent, E., "The Effects of Synthesized Voice Accents on User Perceptions of Robots," International Journal of Social Robots, vol. 3, no. 3, pp. 253–262, Aug. 2011.

[18] Igic, A., Watson, C.I., Macdonald, B.A, Broadbent, E. , Jayawarden, C.J, and Stafford, R., "Perception of Synthetic Speech with Emotion Modeling Delivered through a Robot Platform: An Initial Investigation with Older Listeners", in The Proceedings of the 13th Australasian International Conference on Speech Science and Technology, pp. 189–192, 2010.

[19] Igic, A., Watson, C.I. Teutenberg, J.D, Tamagawa, R., MacDonald, B.A., and Broadbent, E.,"Towards a Flexible Platform for Voice Accent and Expression Selection on Healthcare Robot", in The Proc. 2009 Australasian Language

Technology Workshop. 7, L. A. Pizzato and R. Schwitter (Eds). Sydney, 3 Dec. 2009–4 Dec. 2009, pp. 109–113, 2009.

[20] Bayard D, "The cultural cringe revisited: changes through time in KIWI Attitudes towards accents." In: Bell A, Kuiper K (eds) New Zealand English. Benjamins, Amsterdam, pp 297–324, 1999.

[21] Bayard D, Weatherall A, Gallois C, Pittam J "Pax Americana? Accent attitudinal evaluations in New Zealand", Australia,and America. J Socioling 5:22–49 , 2001.

[22] Cargile A, Giles H "Understanding language attitudes: exploring listener affect and identity." Lang Commun 17:195–217, 1997.

[23] Walters ML, Syrdal DS, Koay KL, Dautenhahn K, te Boekhorst R "Human approach distance to a mechanical-looking robot with different robot voice styles". In: Proc the 17th IEEE international symposium on robot and human interactive communication, Munich, Germany, pp 707–712, 2008.

[24] Goetz J, Kiesler S, Powers A "Matching robot appearance and behavior to tasks to improve human-robot cooperation". In: Proc the 12th IEEE Int. Symp. on Robot and Human Interactive Communication, Millbrae, California, USA, pp 55–60, 2003.

[25] Lister, J., and Tarver, K.,"Effects of age on silent gap discrimination in synthetic speech stimuli," J. Speech Lang. Hear. Res. 47, 257–268, 2004.

[26] Kim S, Frisina RD, Frisina DR. "Effects of age on speech understanding in normal hearing listeners: relationship between the auditory efferent system and speech intelligibility in noise." Speech Communication Vol. 48 855-862, 2006.

[27] Alamsaoutra, D.M., Kohnert, K.J., Munson, B., Reichle, J., "Synthesized speech intelligibility among native speakers and nonnative speakers of English". Augment. Altern. Commun. 22, 258–268. 2006

[28] Jones, C., Berry, L., and Stevens, C.,"Synthesized speech intelligibility and persuasion: Speech rate and non-native listeners," Computer Speech and Language, vol. 21, no. 4, pp. 641–651, Oct. 2007.

[29] Wolters, M., Campbell, P., DePlacido, C., Liddell, A., and Owens, D., "Making Speech Synthesis More Accessible to Older People." 6th ISCA Wrkshp on Speech Synthesis (SSW-6), 2007.

[30] Wolters, M, Campbell, P., DePlacido, C., Liddell, A., and Owens, D., "The effect of hearing loss on the intelligibility of synthetic speech," in Proceedings of the 16th International Congresss of the ICPHS, pp. 673–676, Aug. 2007.

[31] Humes, L.E., Nelson, K.J., Pisoni, D.B., and Lively, S.E., "Effects of Age on Serial Recall of Natural and Synthetic Speech," Journal of Speech and Hearing Research, vol. 36, pp. 634–639, 1993.

[32] Langner A., and Black, A.W., "Using Speech In Noise toImprove Understandability for Elderly Listeners," in Proceedings of ASRU, San Juan, Puerto Rico, 2005.

[33] Roring, R.,W., Hines,F.G., and Charness, N., "Age Differences in Identifying Words in Synthetic Speech," Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 49, no. 1, pp. 25–31, Feb. 2007.

[34] Reynolds, M., Bond, Z., Fucci, D., "Synthetic speech intelligibility: comparison of native and non-native speakers of English". Augmentative and Alternative Communication 12, 32–36. 1996.

[35] Liu, W "Assessing and Improving the Intelligibility of synthetic voices on a healthcare robot" University of Auckland. Unpublished Masters Thesis. 2013

[36] Bradlow A.R., and Bent T., "The clear speech effect for non-native listeners" Journal of the Acoustical Society of America 122 (1), 272-284 2002.

[37] Hongyan, W, and van Heuvan, V.F. "Quantifying the Interlanguage Speech Intellibility Benefit." The proceedings of ICPhS XVI , Saarbrücken, 6-10 Aug , pp 1729-1732, 2007,

[38] Badkar, J., P. Callister & R. Didham. (2009). Aging New Zealand: The Growing Reliance on Migrant Caregivers. Wellington: Victoria University.

[39] R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

[40] Maclagan, M. and Grant, Annabel. (2011) Care of people with Alzheimer's Disease in NZ: Supporting the Telling of Life Stories. Peter Backhaus (ed) Communication in Elderly Care: Cross-Cultural Perspectives London: Continuum. P 62-89.