Int J Soc Robot (2011) 3:253–262
DOI 10.1007/s12369-011-0100-4

ORIGINAL PAPER

The Effects of Synthesized Voice Accents on User Perceptions of Robots

Rie Tamagawa · Catherine I. Watson · I. Han Kuo · Bruce A. MacDonald · Elizabeth Broadbent

Accepted: 13 May 2011 / Published online: 2 June 2011

Abstract Human voice accents have been shown to affect people's perceptions of the speaker, but little research has looked at how synthesized voice accents affect perceptions of robots. This research investigated people's perceptions of three synthesized voice accents. Three male robot voices were generated: British (UK), American (US), and New Zealand (NZ). In study one, twenty adults listened through headphones to a recorded script repeated in the three different accents, rated the nationality, roboticness, and overall impression of each voice, and chose their preferred accent. Study two used these voices on a healthcare robot to investigate the influence of accent on user perceptions of the robot. Ninety-one individuals were randomized to one of three conditions. In each condition they interacted with a healthcare robot that assisted with blood pressure measurement but the conditions differed in the accent the robot spoke with. In study one, each accent was correctly identified. There was no difference in impression ratings of each voice, but the US accent was rated as more robotic than the NZ accent, and the UK accent was preferred to the US accent. Study two showed that people randomized to the NZ accent had more positive feelings towards the robot and rated the robot's overall performance as higher compared to the robot with the US voice. These results suggest that the employment of a less robotic voice with a local accent may positively affect user perceptions of robots.

R. Tamagawa · E. Broadbent (✉)
Department of Psychological Medicine, The University of Auckland, Auckland, New Zealand
e-mail: e.broadbent@auckland.ac.nz

C.I. Watson · I.H. Kuo · B.A. MacDonald
Department of Electronic and Computer Engineering, The University of Auckland, Auckland, New Zealand

Keywords Human-robot interaction · Acceptability · Robot · Voice · Perception

1 Introduction

1.1 Voice in Socially Interactive Technologies

Speech is a primary tool in human communication. Vocal characteristics within speech enable humans to identify various socially relevant cues [32] including gender [29], emotions [2, 39], and social as well as racial background [31]. With the development of socially interactive technology, humans now interact with speech from non-human origins, for example, from assistive robots. This change in communication patterns poses the question of how humans, who are attuned to sustaining social relationships through human voice, respond to speech of non-human origin [31]. One angle from which to approach this question is to find out whether humans respond to social cues within a non-human voice. This study investigates how people perceive synthesized robot voices with different English accents.

1.2 Robot Voice

There is preliminary evidence that voices in socially interactive technologies are important to impressions formed and acceptance of the interface [20, 24, 30, 31, 40]. However, to date the influence of voice has received less attention in robotics research compared with the influence of its appearance. In fact, voice has sometimes been a shortcoming in previous human-robot studies. For instance, a study with college students in USA found that the robot's speech was difficult to understand for non-native English speakers and

therefore they were excluded from the study [20]. In our previous study with a healthcare robot, 60% of the participants reported that they did not like the voice and found the voice too monotonic [24]. The voice was one of the aspects of the robot frequently reported as needing improvement.

Only a handful of work has systematically investigated the potential impact of different types of synthetic voices, let alone robot voices, on user perceptions of the speaker. The results of these studies indicate that people find synthetic speech less trustworthy (in the scenario of a telephone campaign to fundraise) [38], less lively and more squeaky [30] than human speech. Gender has also been shown to influence user perceptions of the speaker. For example, when participants were presented with a persuasive argument, the male synthetic voice was rated as more powerful than the female voice [30]. Goetz and colleagues [20] reported that coherence between the tasks a robot performs and its appearance, including instructive speech, can impact user compliance. Moreover, a recent study showed that voice type could actually influence proximity of human-robot interactions. Walters and colleagues [40] found that when a mechanical looking robot employed either a human male voice, human female voice, synthesized neutral gender voice, or the experimenter's own voice, participants approached significantly closer to the robots with the human voices compared with the synthesized voices. Several studies have investigated how the quality of robotic voice can be improved in terms of pitch and expressiveness [9, 13, 23, 26, 37], but little is known about how the accent of the voice impacts on users' perceptions of a robot.

1.3 Perceptions of Voice Accent

There has been longstanding interest in understanding people's attitudes towards different accents [15]. People infer many different things about a person from the way they speak, from likeability, and trustworthiness right through to ambition, and social standing [6–8, 15, 27]. If people have negative perceptions of a voice, it can prevent the communicative intent of the spoken word being properly understood [15].

Each English speaking country or region has its own particular spoken accent of the English language, although the popular media exposes populations to a number of different accents. In New Zealand (NZ), since the 1960s United States (US) television programmes have been screened and have formed over half of the viewing for many years, while New Zealand-made programmes provide around 20% of the content [7]. However, it is not certain whether this amount of exposure to foreign accents influences the attitude towards one's own and other English accents. Interestingly, studies during the 1970s reported that New Zealanders rated British accents highly pleasant and socially desirable relative to other foreign accents and the NZ accent [6, 7]. Subsequent studies through the 1980s to late 1990s with New Zealanders [7, 22] as well as Australians [5] have shown that the British accent has continued to be rated highest in charisma, power, competence and pleasantness, while the North American accent often rated a close second [7, 22]. Similarly, a more recent study in Singapore using a virtual help desk assistant showed that UK accent was preferred over a Singaporean accent [33]. Thus, there seems to be a consistent trend for people to perceive a "BBC" English accent as prestigious relative to other accents. This may be a particular trend in places such as New Zealand, Australia and Singapore, which were once British colonies.

Several studies suggest that culture may be an important variable in preferences for voice accents. For example, Nass and Brave [31] found that first generation Korean Americans rated a Korean-accented voice agent more positively than an Australian-accented voice agent. On the other hand, Caucasian Americans preferred the Australian voice to the Korean voice, because the Australian culture was perceived as closer than the Korean culture to their own American culture. In a study of multimedia learning with American students, an American accent was preferred over a Russian accent, and resulted in better learning outcomes [28]. Moreover, in the New Zealand studies described above, the NZ accent was rated highest on acceptableness amongst all other accents [7, 22]. Other work with synthetic speech has shown a local accent was preferred for "fun" applications such as a talking clock or a game, whereas a prestigious accent was preferred for applications in the health domain [35]. Thus, a sense of familiarity or similarity seems to influence peoples' preference for accents [31] as well as type of application. The use of social cues in the voice, such as the user's own type of accent and appropriate application tone, may foster favorable interactions between socially interactive technology and humans.

We know little about whether attitudes towards foreign and own English accents translate to the context of human-robot interactions. More specifically, it is not known whether a robot's accent can influence a human user's feelings towards the robot and ratings of its performance.

1.4 Aims of this Research

In our previous work, we examined how adult New Zealanders evaluated a healthcare robot that gave instructions for using a device for blood pressure (BP) measurement with an American-accented voice [24]. Many participants wrote that the voice was too robotic and/or monotone. As a continuum of our project to develop a healthcare robot in New Zealand, the voice needed to be improved. In this study, we specifically chose two synthesized voices that sounded less robotic

to our own ears, and investigated whether they could result in improved ratings of the robot overall. In study one, we examined how New Zealand adults evaluated the three different synthesized voices without any robot present: the original United States (US) voice, a New Zealand voice (NZ), and a United Kingdom (UK) voice. We investigated whether participants could identify each accent, whether the voices differed in perceived roboticness and which was the preferred voice. The following hypotheses were tested in study one.

H1a: the original US voice would be rated more robotic-sounding than the other two voices.
H1b: the UK voice and NZ voices would be preferred to the US voice.

In study two, we examined whether employing the different voices on a robot could affect participants' feelings towards the robot and their evaluation of the robot's performance. For study two, we tested the following primary hypothesis:

H2: a robot with a UK or NZ voice would be rated as having a better performance compared with a robot with a US voice.

2 Study One

This study aimed to determine whether three synthesized voice accents differed in how robotic they sounded, and which voice people preferred.

2.1 Method

2.1.1 Participants

Twenty individuals working or studying in the Department of Psychological Medicine, the University of Auckland, New Zealand, were approached and asked if they would be willing to take part in a study assessing different synthesized robot voices. All twenty people approached agreed to participate. Eighteen participants were post-graduate students and two were staff members. There were six males and 14 females. The mean age was 31.95 ($sd = 11.65$). On average, the participants had lived in NZ for 20.87 years ($sd = 12.26$), and eight (40%) were born in NZ.

2.1.2 Procedure

Each participant was asked to listen to three different synthesized voices (NZ, US, and UK) reading the same script, one at a time through headphones attached to an MP3 device. The script was the same for both studies 1 and 2 and included a general greeting; "Hello, my name is Charles", and instructions of how to operate a blood pressure device, starting "Please pick up the blood pressure cuff…" The three voices were presented in random order. Following each voice, the participants were asked to complete a short questionnaire.

The synthetic speech is generated within the Festival Speech Synthesis framework (http://www.cstr.ed.ac.uk/projects/festival/). Within this system, input text is transformed into a speech output. There are a number of voices available within the Festival system. In this study we used two of these voices: KAL, which is an American male voice, and RAB, which is a British male voice. The third voice was a New Zealand English voice recently developed by the 2nd author [42]. All three voices were developed using the same diphone concatenation method [11, 12]. Diphones were created from nonsense phrases that were recorded at a sampling rate of 16 kHz with a bit size of 16 bits. Diphones are two sequential sounds, for example the word cat is made up of four diphones, silence-/k/, /k/-/ae/, /ae-t/, and /t/-silence. There are just over 2100 diphones in English. The diphones are extracted from phrases that were recorded at a sampling rate of 16 kHz with a bit size of 16 bits. The phrases were constructed specifically for developing the synthetic voice [11]. The output synthesized speech had the same sample rate and bit size as the recorded speech. All three voices used the same prosody model, and duration models [11] but each had their own separate pronunciation dictionary, for which the pronunciations for each word in the lexicon were specific for the particular English accent. We ensured the sampling rate and bit-size for all three voices was the same. Therefore, the output synthesized speech was of the same quality in terms of how easy the speech was to listen to. Whilst the pronunciation dictionaries for the KAL and RAB voices came with the Festival system, we had to generate our own pronunciation dictionary for the New Zealand voice using the process as outlined [17].

2.1.3 Measures

Participants were asked to indicate the nationality of each voice from a list of nine different nationalities: New Zealand, Australian, American, Canadian, British, Irish, Asian, South African, or other/non-identifiable.

A voice attribute scale [7] was used to assess perceived voice impression. This scale has been used in previous research to rate human voices. We modified the scale slightly to make it more appropriate for rating a robot's voice. Participants were asked to rate the voice on a 5 point scale from strongly positive '1' to strongly negative '5' for the following 11 characteristics; pleasantness, reliability, ambition, leadership, intelligibility, knowledgeableness, self-confidence, intelligence, likability, acceptability, and authoritativeness. A high internal consistency indicated by the

Cronbach's alpha of .90, and the mean inter-item correlations of .51, suggest the current scale provides an overall measure of how positively the voice is perceived. The ratings of 11 items were summed to give a voice impression score. A lower score indicates better voice impression. [The modifications we made to the original scale were: we removed the attributes of "estimated income", "perceived socio-economic class", "education level" and "sense of humor" because robots do not live in society, go to school, receive an income, or have socio-economic class, so we felt participants may have particular difficulty with these items. Furthermore, research has shown that people have higher exercise adherence in response to a serious robot than a fun robot and this suggests a serious robot is more appropriate than a humorous one for health applications [20]. These attributes were replaced with "authoritativeness", "intelligibility" and "knowledgeableness" because research shows that perceived authoritativeness and knowledgeableness of the source are important in healthcare [21, 25] and intelligibility is important for people to be able to follow the robot's instructions].

Participants were asked to rate "how robotic did the voice sound to you?" on a 10-point scale, ranging from very natural, human-like '1' to very mechanical, robot-like '10'. At the end of the session, participants were also asked which voice they liked the most.

2.1.4 *Statistical Analyses*

All statistical analyses were performed with SPSS version 15. A Chi-square test was conducted to examine whether participants correctly detected the accent of each robot voice. A one-way analysis of variance (ANOVA) was used to compare the ratings of voice impression, and perceived roboticness between the three accents. Three post-hoc pairwise comparisons (NZ vs. US, US vs. UK, UK vs. NZ) were followed when the omnibus effect was significant. In order to control Type I error, the significant level for each pairwise comparison was set as .016, using a Bonferroni correction. Preferred accent was also examined using a Chi-square test.

2.2 Results

2.2.1 *Identifying the Nationality of the Voice*

As shown in Table 1, approximately half of the participants correctly identified the accent of each voice. Results of Chi-square analyses were significant for each voice: UK, χ^2 $(5, N = 20) = 17.2$, $p = .004$; US, χ^2 $(5, N = 20) = 13.6$, $p = .018$; NZ, χ^2 $(3, N = 20) = 17.2$, $p = .001$. Thus, for each accent, significantly more participants chose the correct nationality of the voice compared to the other nationalities.

2.2.2 *Voice Impression, Roboticness, and Preferred Accent*

Table 2 shows means and standard deviations for the voice impression score and the ratings of roboticness for the three voices, as well as the proportion of the participants who preferred each accent. Results of ANOVA showed no significant differences in the voice impression scores (Table 2). However, the rated roboticness of the three voices were significantly different. Post-hoc analyses revealed that the NZ accent was rated significantly less robotic relative to the US accent (mean difference = 1.80, $p = .002$). No other post-hoc analyses showed a significant difference. Results of a

Table 1 Distributions of identified nationalities for each synthesized voice in study one

	UK ($n = 20$) *n* (%)	US ($n = 20$) *n* (%)	NZ ($n = 20$) *n* (%)
New Zealand	3 (15)	0	13 (65)
Australia	0	0	2 (10)
American	0	9 (45)	0
Canadian	0	1 (5)	0
UK	10 (50)	0	2 (10)
Irish	2 (10)	4 (20)	0
Asian	1 (5)	2 (10)	0
South African	3 (15)	1 (5)	3 (15)
Other	1 (5) (Robotic)	3 (15) (Arabic, unidentifiable)	0

Table 2 Means and SD, and statistical results of the voice impression score and the roboticness score for the three accents, and the number of people who preferred each voice in study one

	UK ($n = 20$) *M* (SD)	US ($n = 20$) *M* (SD)	NZ ($n = 20$) *M* (SD)	Statistics $F(2, 57)$	p
Impression[a]	27.45 (6.91)	31.85 (6.04)	31.75 (7.82)	2.61	.08
Roboticness	7.25 (1.83)	8.30 (1.22)	6.50 (2.19)	5.94	.009
	n (%)	*n* (%)	*n* (%)	χ^2 $(2, N = 20)$	p
Preferred	11 (55)	2 (10)	7 (35)	6.10	.04

[a]Lower score indicates better impression

Chi-square test showed the number of people who preferred each voice varied significantly between the robot voices, χ^2 $(2, N = 20) = 6.10$, $p = .047$. Significantly more people preferred the UK accent to the US accent, χ^2 $(1, N = 13) = 6.23$, $p = .013$. No other pairwise comparisons revealed a significant difference, $p > .05$.

2.3 Discussion Study One

The NZ accent was rated significantly less robotic than the US accent, while the difference in roboticness between the UK and other voices was not significant. These findings partially support H1a, that the US voice would be rated more robotic-sounding. The lower roboticness rating of the NZ voice may be partly due to similarity with the participants' own culture. Many participants were able to correctly identify each accent and would have been most familiar with the NZ accent. Similarity of culture has been shown to be important to ratings of voices in previous work [31]. With its colonial past, the New Zealand culture is closer to the UK culture than the US culture, which may explain why this difference in voice ratings was significant between the NZ and US voices but not between the UK and NZ voices.

Significantly more participants preferred the UK accent than the US accent. These findings partially support hypothesis H1b, that the UK and NZ voices would be preferred to the US voice. These results are in line with previous research in New Zealand that has shown the UK accent is considered more pleasant and socially desirable than both the New Zealand accent and other foreign accents [6, 7]. This preference for the UK accent over other foreign English accents seems to still be a trend in New Zealanders. The non-significant differences in preference between the NZ and other voices may mean that the NZ voice is becoming more liked, although these differences may have reached significance had the sample size been larger.

All three voices were generated in the same way (see Sect. 2.1.2), however it is possible that the New Zealand voice recording conditions and equipment for recording the nonsense phrases were of a higher standard than the two other voices, which were created over 12 years ago. Further, although much of the voice generation process is automated, all the automated processes were hand checked, and considerable care and months were put into developing the voice. The KAL voice by contrast took two days to construct [12]. However, this cannot explain the observed preference differences between the UK and US voices, nor why there were differences in roboticness ratings between the NZ and US voice but not between the NZ and UK voices.

We chose to do this first study on the voice alone, without a physical embodiment of a robot, to establish whether the accents did indeed differ simply as synthesized voices. The study established that there were some differences in perceived roboticness and overall preference between accents of synthesized voice. We were then interested in whether these different accents could influence user ratings when they were spoken by a robot. The second study investigated whether altering the accent of a robot would influence people's feelings towards and ratings of the robot in a between subjects design.

3 Study Two

The primary aim of this study was to investigate whether a healthcare robot's voice accent affected people's ratings of the robot and feelings towards it.

3.1 Method

3.1.1 Participants

A total of 92 individuals were recruited through electronic mail advertisements and notice boards within the University of Auckland, Auckland, NZ. One participant could not complete the whole experimental session due to a fire alarm at the campus. A total of 91 participants completed an individual experiment and pre- and post-experimental assessments.

3.1.2 Procedure

The study was approved by the University of Auckland Human Participants Ethics Committee. Individuals who responded to the advertisements were asked to make an individual appointment for a 45-minute study session. On arrival, each individual was asked to complete a consent form followed by a baseline questionnaire that assessed expectations and feelings towards a healthcare robot. Then, each participant was invited into a separate room and was seated in front of a table. At this point, participants were randomized to one of the three experimental conditions, in which the healthcare robot utilized either a UK, US, or NZ accent. The experiment followed a standard procedure wherein each participant interacted with the healthcare robot (Charles, see Fig. 1), who assisted with blood pressure (B.P.) measurements as per previous studies [24].

The robot moved from its standby position to a pre-set place near a table in front of a participant. The robot moved close enough to the table, so participants were able to reach the B.P. device on the robot. The robot briefly introduced itself and then provided instructions for the B.P. measurement procedure. In order to ensure proper B.P. measurements, the robot provided step by step instructions vocally while presenting pictures on the screen for how to position the cuff on the arm, which button to press to start the measurement, and how the cuff would inflate and deflate. Following the

Fig. 1 The robot used in the study

B.P. measurement, the robot vocally provided the results, informed the user that the session was completed, and moved back to its start position. This scenario involved mostly one-way communication from the robot to participants on how to measure B.P. and their blood pressure results. Participants were simply required to follow the instructions, properly position the cuff, and press the start button when they were ready. Because the procedure was always the same, it was easy to ensure consistent performance from the robot.

The session involved two B.P. and heart rate measurements; one taken by the robot and one by the research fellow. The order was alternated to avoid an order effect. Except for the accent, all other robot features including the face and instructions, were exactly the same in all three conditions. Following the experiment, each participant was asked to complete a post-task questionnaire, and was offered a movie voucher. The participants were unaware that the purpose of the study was to compare how people reacted to different voice accents. They were told simply that study was to see how participants rated the robot but not told that participants were randomized to different robot voice conditions. This study required the participant fill in the first questionnaire before encountering the robot (Charles), and move to a separate room for a short interaction with Charles. A research fellow was present in the room to ensure consistent experimental procedure for all participants, and a graduate engineering student was available throughout for the maintenance of the robot such as charging the battery and software performance.

3.1.3 Measures

The baseline questionnaire was used to gather information about demographic characteristics, and feelings towards a healthcare robot using the Positive and Negative Affect Schedule (PANAS) [41]. The PANAS includes 20 adjectives and is divided into two subscales. Ten items assess positive affect and ten items measure negative affect. Participants were asked to rate how they felt towards using a healthcare robot on a 5-point scale from not at all '1' to Extremely '5'. Each subscale was scored separately, and a high score indicates a greater correspondent emotionality.

Following the brief session with the healthcare robot (approximately 5 minutes), participants were asked to fill in the PANAS twice, once with instructions to rate how they had felt while they were interacting with the robot and the second time with instructions to rate how they felt now, after the interaction. We asked people to fill in the PANAS twice because we were interested in whether feelings changed from during to after the interaction. However, participants' scores on the PANAS were not significantly different between these two repetitions of the scale, so the two scores were averaged.

The post-task questionnaire included the same voice impression scale as in study one. Participants were asked to indicate the nationality of the robot based on the accent of the voice from nine alternatives: New Zealand, Australian, American, Canadian, British, Irish, Asian, South African, or other.

The quality of interaction was defined as the degree to which people felt pleasant and comfortable in a social encounter [10]. This was examined using the summed combination of six items from the Social Interaction Scale, which was validated in social psychology to study the quality of one-on-one human interactions [10]. The items were 'how much was the interaction satisfying?', 'how much did you enjoy the interaction?', 'how smooth, natural, and relaxed was the interaction?', 'how forced, strained, or awkward was the interaction?' (reverse coded), 'how much would you like to interact with the robot again?', and 'how much was the interaction pleasant?' rated from 0 to 8 (Cronbach's alpha = .91, Mean inter-item correlation = .63). This scale has been used in previous work to measure the quality of human robot interactions [14, 24]. Factor analyses were performed, and the results of principal component analyses showed only one component for the six items. Participants were asked to indicate how much they liked the features of the robot, e.g., face and voice, and how they would rate the overall performance of the robot on a single item using a 8-point scale ranging from very poor '1' to excellent '8'.

3.1.4 Statistical Analysis

Both, age and the number of years living in NZ met the assumptions of normality. Preliminary analyses were performed to examine pre-existing differences in age, gender,

Table 3 Demographic characteristics of participants in study two

		M (SD)
Age		31.38 (10.31)
Living in New Zealand		19.94 (15.46)
		n (%)
Gender	Male	29 (31.5)
	Female	63 (68.5)
Ethnicity	NZ European	43 (46.7)
	European	10 (10.9)
	Asian	15 (16.3)
	Indian	9 (9.8)
	Australian	2 (2.2)
	American	2 (2.2)
	Other	11 (12.1)

ethnicity background, and the number of years living in NZ between the three voice conditions. Potential differences in age and the number of years living in NZ were examined by using ANOVA. Chi-square tests were used to compare the three conditions for gender and ethnicity distributions. Differences in the outcome measures were examined between the three voice conditions. This study included results for participants' subjective reports on positive and negative feelings towards the robot, and impression ratings for the robot interaction, and robot's voice. For all the outcome measures, one-way analyses of covariance (ANCOVA) were performed with gender as a covariate. For positive and negative feelings, baseline values were taken into account in the analyses to ensure post-task differences were not a result of mood states prior to meeting the robot. When those omnibus effects were significant, three pairwise contrasts (NZ vs. UK, NZ vs. US, and US vs. UK) were followed to examine the differences between the three voices. Bonfferoni methods were applied to control Type I error, and a significant level was set as .016 for each pairwise comparison. The data for negative affect (NA) was skewed, therefore, a square-root transformation was applied before analyses.

3.2 Results

3.2.1 Participants Characteristics

Table 3 summarizes demographic characteristics of the participants. The majority of participants were from the NZ European ethnic group. Other ethnic groups included South African, Canadian, Sri-Lankan, Brazilian, Muslim, and mixed (e.g. half Maori and half NZ European).

Table 4 Participants' identification of nationality of the robot voice in the three conditions in study two

Answers	UK ($n = 31$) *n* (%)	US ($n = 29$) *n* (%)	NZ ($n = 32$) *n* (%)
New Zealand	2 (6.5)	4 (13.8)	17 (53.1)
Australia	0	1 (3.4)	1 (3.1)
American	5 (16.1)	14 (48.3)	3 (9.4)
Canadian	2 (6.5)	2 (6.9)	2 (6.3)
UK	19 (61.3)	2 (6.9)	3 (9.4)
Irish	0	1 (3.4)	0
Asian	2 (6.5)	1 (3.4)	0
South African	0	2 (6.9)	2 (6.3)
Other	1 (3.2)	2 (6.9)	4 (12.5)

Results of preliminary analyses for pre-existing differences between the three voice conditions showed a significantly different distribution of gender, χ^2 (2, $N = 92$) = 7.18, $p = .028$, but not ethnic groups, χ^2 (12, $N = 92$) = 13.29, $p = .35$. The distribution of males was lower in the UK voice condition (17.2%) relative to the other two conditions (US: 48.3%, NZ: 34.5%). There were no pre-existing baseline differences in age, $F(2, 89) = .55$, $p = .58$, and the number of years living in NZ, $F(2, 89) = .06$, $p = .94$, between the three conditions. Due to the different distribution of gender across the three groups, gender was entered as a covariate for the rest of the analyses.

3.2.2 Identifying the Nationality of the Voice

Table 4 shows the proportions of nationalities identified for each voice. Results of a Chi-square test within each voice condition were significant: UK, χ^2 (6, $N = 31$) = 46.23, $p < .001$, US, χ^2 (8, $N = 29$) = 42.69, $p < .001$, NZ, χ^2 (6, $N = 32$) = 40.63, $p < .001$. As shown in Table 4, the proportion of participants who correctly identified each accent was significantly greater than those who chose the other accents. Those who identified the voice as other, reported the accent was unidentifiable or robotic.

3.2.3 Positive and Negative Feelings Towards Robots

Table 5 shows means and standard deviations for the PANAS. The results of ANCOVA showed a significant difference in the level of positive emotions between the three voice conditions. Follow-up contrasts revealed that those who were in NZ accent condition reported significantly greater positive emotions than did those in the US accent condition ($p = .014$). No other pairwise comparisons reached the significant difference levels. No differences were found for negative feelings towards a healthcare robot between the three voice conditions.

Table 5 Means and standard deviations for the PANAS in the three voice conditions in study two

	UK ($n = 31$)	US ($n = 29$)	NZ ($n = 32$)	Statistics	
	M (SD)	M (SD)	M (SD)	$F(2, 87)$	p
Positive affect	30.77 (9.55)	26.07 (7.13)	29.58 (8.96)	3.18	.04
Negative affect[a]	11.00 (2.50)	11.50 (3.25)	10.75 (2.00)	2.70	.07

[a]The Negative affect data were skewed, therefore, median and inter-quartile ranges were included in the table

3.2.4 *Impression of Robot Voice, and Quality of the Interaction*

There was no difference in the total score of voice impression ratings, $F(2, 88) = .17,\ p = .84$, or the quality ratings for the interaction, $F(2, 88) = 1.14,\ p = .32$, between the three voice conditions. The three conditions reported significantly different levels of overall performance of the robot, $F(2, 88) = 3.51,\ p = .03$. Those in the NZ accent condition reported significantly better overall robot performance (mean 6.25, SD 2.28) relative to those who were in the US accent condition (mean 5.31, SD 2.47) ($p = .01$). There were no significant comparisons in performance with the UK version (mean 5.84, SD 2.37).

No significant difference emerged for the ratings of the extent to which people liked the face, $F(2, 88) = .61,\ p = .54$, the voice of the robot, $F(2, 88) = .92,\ p = .40$, or the clarity of the voice, $F(2, 88) = 1.47,\ p = .23$.

4 Discussion and Conclusions

The second study aimed to investigate whether a less robotic voice could result in higher ratings of a robot using a between-subjects experimental design. The findings showed that people who were randomized to the robot with the NZ voice, which was rated as less robotic in study one, had more positive emotions during the interaction than people randomized to the robot with a US accent. Furthermore, the overall performance of the robot with the NZ accent was rated as higher than that of the robot with the US accent. This finding is consistent with previous work that demonstrated that a web-based agent speaking with the accent of the participant's own ethnic background was perceived as more positive relative to a foreign accent in terms of overall agent quality [31]. In keeping with findings on human speech [6–8, 15, 27], our results suggest that perceptions of the robot were altered by its accent.

The results partially support hypothesis H2, that a robot with a UK or NZ voice would be rated as having a better performance compared with a robot with a US voice. In study 1, the UK voice was preferred to the US voice, but there were no significant differences in ratings of the robot between these two accents in study 2. There was only a difference in ratings of the robot between the NZ and US voices, and the first study showed that these voices differed in roboticness.

These results can have two interpretations. The first is that a less robotic voice makes people think the robot performs better and they feel more positive towards it. The second interpretation is that having a local accent on the robot increases perceptions of performance and makes people feel more positive towards it (although the fact that there were no significant differences between ratings of the robots with the UK and NZ voices weakens this interpretation).

It is likely that both these interpretations have some validity, and to tease these effects out further, the study would have to be repeated in the UK, the US or in a neutral country, to compare results. It may be that residents of the US prefer the US voice and/or rate it as less robotic. Each accent elicits cultural stereotypes of the region [18, 19], thus we cannot rule out the possibility that people's preferences for certain accents are influenced by stereotypes attached to a given region or country, at both personal and global levels. It is possible that the contents of speech (instructions for B.P. measurement) might have influenced the results, as previous studies indicate that people have different preferences for accents depending on the application type [20]. Further research needs to investigate these possibilities.

It is interesting to consider how generalizable these results might be to non-native speakers of English, who are likely to be less able to distinguish between different English accent types, but may find stronger accents harder to understand.

For synthetic speech alone there is evidence that non-native speakers find synthetic speech harder to comprehend than human speech [1]. The implications for robot makers are to consider making the voice more natural and human-like to increase acceptance. This is possibly in contrast to how human-like the robot looks, where previous research suggests that people prefer a less humanlike looking robot [3, 16, 34, 36]. Vocal characteristics, such as accent have been found to be more important than mere physical appearance in eliciting positive responses to socially interactive technology [31]. Robot designers might consider customizing the accent of the robot to the country in which it is employed, or providing users with a number of voice op-

tions. Speech interfaces may be useful in assistive robotics for people with mobility impairments [4].

In conclusion, this study has shown that people are influenced by the accents of synthesized voices when they rate the performance of robots, and these accents also influence people's experience of positive feelings. The importance of speech accent may be extended from human voices in human-human interactions to synthetic robot voices in human-robot interactions.

Acknowledgement This research was funded by a University of Auckland Grant to Elizabeth Broadbent.

References

1. Alamsaputra DM, Kohnert KJ, Munson B, Reichle J (2006) Synthesized speech intelligibility among native speakers and non-native speakers of English. Augment Altern Commun 22:258–268
2. Aronovitch CD (1976) The voice of personality: Stereotyped judgment and their relation to voice quality and sex of speaker. J Soc Psychol 99:207–220
3. Arras KO, Cerqui D (2005) Do we want to share our lives and bodies with robots? A 2000 people survey. Technical report, Autonomous Systems Lab, Swiss Federal Institute of Technology Lausanne
4. Atrash A, Kaplow R, Villemure J, West R, Yamani H, Pineau J (2009) Development and validation of a robust speech interface for improved human-robot interaction. Int J Soc Robot 1:345–356
5. Ball P (1983) Stereotypes of anglo-saxon and non anglo-saxon accents: some exploratory Australian studies with the matched guise technique. Lang Sci 5:163–183
6. Bayard D (1995) Kiwitalk: sociolinguistics and New Zealand society. Dunmore Press, Palmerston North
7. Bayard D (1999) 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
8. Bayard D, Weatherall A, Gallois C, Pittam J (2001) Pax Americana? Accent attitudinal evaluations in New Zealand, Australia, and America. J Socioling 5:22–49
9. Bennewitz M, Faber F, Joho D, Behnke S (2007) Fritz—a humanoid communication robot. In: Proc IEEE international workshop of robot and human interactive communication (ROMAN), Jeju Island, Korea, pp 1072–1077
10. Berry DS, Hansen JS (1996) Positive affect, negative affect, and social interaction. J Pers Soc Psychol 71:796–809
11. Black AW, Lenzo KA (2007) Building synthetic voices. http://festvox.org/festvox/festvox_toc.html
12. Black AW, Taylor P, Caley R (1999) The festival speech synthesis system. http://www.cstr.ed.ac.uk/projects/festival/
13. Breazeal C (2001) Emotive qualities in robot speech. In: Proc the 2001 IEEE/RSJ international conference on intelligent robots and systems, Maui, pp 1388–1394
14. Broadbent E, MacDonald BA, Jago L, Juergens M, Mazharullah O (2007) Human reactions to good and bad robots. In: Proc IEEE/RSJ international conference on intelligent robots and systems IROS, pp 3703–3708
15. Cargile A, Giles H (1997) Understanding language attitudes: exploring listener affect and identity. Lang Commun 17:195–217
16. Cesta A, Cortellessa G, Giuliani MV, Pecora F, Scopelliti M, Tiberio L (2007) Psychological implications of domestic assistive technology for the elderly. Psychol J 5:229–252
17. Fitt S (2000) Documentation and user guide to Unisyn Lexicon and Post-Lexical rules. Technical report, Centre for Speech Technology Research, University of Edinburgh
18. Giles H (1970) Evaluative reactions to accents. Educ Rev 22:211–227
19. Giles H, Williams A, Mackie DM, Rosselli F (1995) Reactions to Anglo- and Hispanic-American-accented speakers: Affect, identity, persuasion, and the English-only controversy. Lang Commun 15:107–120
20. Goetz J, Kiesler S, Powers A (2003) Matching robot appearance and behavior to tasks to improve human-robot cooperation. In: Proc the 12th IEEE international symposium on robot and human interactive communication, Millbrae, California, USA, pp 55–60
21. Hall JA, Roter DL, Rand CS (1981) Communication of affect between patient and physician. J Health Soc Behav 22:18–30
22. Huygens I, Vaughan GM (1983) Language attitudes, ethnicity and social class in New Zealand. J Multiling Multicult Dev 4:207–223
23. Igic A, Watson CI, Teutenberg J, Tamagawa R, Macdonald BA, Broadbent E (2009) Towards a flexible platform for voice accent and expression selection on a Healthcare Robot. In: Proc the 2009 Australasian language technology workshop, Sydney, pp 109–113
24. Kuo IH, Rabindran J, Broadbent E, Lee YI, Kerse N, Stafford R, MacDonald BA (2009) Age and gender factors in user acceptance of healthcare robots. In: Proc the 18th IEEE international symposium on robot and human interactive communication, Toyama, Japan, pp 214–219
25. LeBaron S, Reyher J, Stack JM (1985) Paternalistic vs egalitarian physician styles: The treatment of patients in crisis. J Fam Pract 21:56–62
26. Li X, Watson CI, Igic A, Macdonald BA (2009) Expressive speech for a virtual talking head. In: Australasian conference on robotics and automation, Sydney
27. Luhman R (1990) Appalachian English stereotypes: Language attitudes in Kentucky. Lang Soc 19:331–348
28. Mayer RE, Sobko K, Mautone PD (2003) Social cues in multimedia learning: roles of speaker's voice. J Educ Psychol 95:419–425
29. Mullennix JW, Johnson K, Topcu-Durgun M, Farnsworth LW (1995) The perceptual representation of voice gender. J Acoust Soc Am 98:3080–3095
30. Mullennix JW, Stern SE, Wilson SJ, Dyson C (2003) Social perception of male and female computer synthesized speech. Comput Hum Behav 19:407–424
31. Nass C, Brave S (2005) Wired for speech: how voice activates and advances the human-computer relationship. MIT Press, Cambridge
32. Nass C, Gong L (2000) Social aspects of speech interfaces from an evolutionary perspective: experimental research and design implications. Commun ACM 43(9):36–43
33. Niculescu AI, White GM, Lan SS, Waloejo RU, Kawaguchi Y (2008) Impact of English regional accents on user acceptance of voice user interfaces. In: Proc NordiCHI 2008, vol 358. ACM, New York, pp 523–526
34. Oestreicher L (2007) Cognitive, social, sociable or just socially acceptable robots. In: Proc the 16th IEEE international symposium on robot and human interactive communication (ROMAN), Jeju Island, Korea, pp 558–563
35. Pucher M, Schuchmann G, Fröhlich P (2008) Regionalized text-to-speech systems: Persona design and application scenarios. In: COST action 2102 school, Vietri sul Mare, Italy. Lecture notes in artificial intelligence (LNAI), vol 5398, pp 216–222
36. Robins B, Dautenhahn K, te Boekhorst R, Billard A (2004) Robots as assistive technology—does appearance matter. In: Proc the 13th IEEE international workshop on robot and human interactive communication (ROMAN), Okayama, Japan, pp 277–282
37. Roehling S, MacDonald BA, Watson C (2006) Towards expressive speech synthesis in English on a robot platform. In: Proc the 11th

Australian international conference on speech science and technology, Auckland, New Zealand, pp 130–135
38. Stern SE (2008) Computer-synthesized speech and perceptions of the social influence of disabled users. J Lang Soc Psychol 27:254–265
39. Tusing KJ, Dillard JP (2000) The sound of dominance: vocal precursors of perceived dominance during interpersonal influence. Hum Commun Res 26:148–171
40. Walters ML, Syrdal DS, Koay KL, Dautenhahn K, te Boekhorst R (2008) 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
41. Watson D, Clark LA, Tellegen A (1988) Development and validation of a brief measure of positive and negative affect: the PANAS scales. J Pers Soc Psychol 54:1063–1070
42. Watson CI, Teutenberg J, Thompson L, Roehling S, Igic A (2009) How to build a New Zealand voice. In: NZ linguistic society conference, Palmerston North

Rie Tamagawa received her B.A. in Psychology from the University of Canterbury, Christchurch, New Zealand, and her M.Sc. and Ph.D. in Health Psychology from the University of Auckland, New Zealand. Rie is currently a Post-doctoral research fellow in Psycho-Oncology at the Department of Oncology at the Faculty of Medicine, the University of Calgary in Canada. Her research interests include cross-cultural comparisons of emotions, and emotional reactions involved in human-robot interaction.

Catherine I. Watson received a B.E. (hons) and Ph.D. from the Department of Electrical and Electrical Engineering, University of Canterbury, New Zealand. She spent 8 years at Macquarie University, Sydney, Australia, first in a post doctorate in the Speech Hearing and Language Research Centre in the Department of Linguistics, this morphed then into a joint position in the Macquarie Centre for Cognitive Science and a lectureship in the Department of Electronics. In 2004 she returned to New Zealand and took up a position in the Department of Electrical and Computer Engineering. Catherine's research interests include modelling and analysing speech production in humans and machines, and human-machine interaction.

I. Han Kuo received a B.E. (1st class) in computer systems engineering from the University of Auckland, where he is currently studying for a Ph.D. in Human-robot interaction (HRI) and healthcare robotics. Research interests include HRI design methodology, psychological theories on HRI, and development of multi-modal (mixed-initiative) communication for robot service applications.

Bruce A. MacDonald received a B.E. (1st class) and Ph.D. in the Electrical Engineering department of the University of Canterbury, Christchurch, New Zealand. He spent ten years in the Computer Science department of the University of Calgary in Canada then returned to New Zealand in 1995, joining the Department of Electrical and Computer Engineering Department at the University of Auckland. He is director of the Robotics Laboratory. The long term goal is to design intelligent robotic assistants for that improve the quality of peoples' lives. His research interests include human robot interaction and robot programming systems, with applications in areas such as healthcare and agriculture.

Elizabeth Broadbent received her B.E. (Hons) in Electrical and Electronic Engineering from the University of Canterbury, New Zealand, and her M.Sc. and Ph.D. degrees in health psychology from the University of Auckland. Elizabeth is currently a Senior Lecturer in Health Psychology at the Faculty of Medical and Health Sciences, the University of Auckland, where she is the Director of the Health Psychology PGDipSci and M.Sc. programme. Her research interests include human robot interaction, with a particular interest in emotional reactions to robots and perceptions of robots.