Raising Language Awareness Using Digital Media
– methods for revealing linguistic stereotyping

Mats Deutschmann, Department of Language Studies, Umeå University, Sweden.
Anders Steinvall, Department of Language Studies, Umeå University, Sweden.
Anna Lagerström, Department of Language Studies, Umeå University, Sweden.

Abstract
Language is closely connected with identity in a number of distinctive ways. Not only is language closely linked to nationality, but we also make conscious and unconscious associations of language output and group identity (class, gender, generation, ethnicity) whenever we meet someone. It has long been demonstrated that individuals are judged in terms of intellect and other character traits on the basis of their voice quality, intonation and accent, something which will affect 'identification', the ongoing, interactive process of identity construction that takes place during all human interaction (Cavallaro & Chin, 2009; Fuertes et al. 2012; Deutschmann, Steinvall & Lagerström, 2011)

This chapter gives an account of methods for raising awareness of language issues and identity (stereotyping for example) using digital media. We shall discuss how working in a virtual world environment with digital representations of speakers (avatars), we used avatar manipulations and so-called voice morphing tools with the aim of expose students to different disguised 'versions' of the same speaker (by manipulating gender for example). We were able to show that their judgement of the same person differed greatly depending on how the latter had been digitally manipulated, and we used this as a starting point for discussions on language and identity. The method has great implications for language-awareness-raising activities, particularly in vocational training of professionals who work with human contacts on a daily basis (teachers, police, nurses and doctors, for example).

Mats Deutschmann is Associate Professor in English Didactics and Linguistics at the Department of Language Studies, Umeå University, Sweden. He has published extensively in the fields of sociolinguistics and CALL (computer assisted language learning) with a focus on digital media for innovative design in language learning and language-awareness-raising activities.

Anders Steinvall is Senior Lecturer in English Linguistics at the Department of Language Studies, Umeå University, Sweden. He is an experienced educator, with more than 15 years of practice teaching online. In his every-day teaching, Steinvall frequently makes use of ICT as a tool for stimulating discussions and debriefings in which students reach new insights.

Anna Lagerström is a freelance translator who took part in this study as a linguistics student.

1. Introduction

Whether we are aware of it or not, language is at the heart of the mechanisms leading to stereotyping and inequality. It is one of the major factors that we evaluate when we meet others, and it has long been demonstrated that individuals are judged in terms of intellect and other character traits on the basis of their language output (e.g. Cavallaro & Chin, 2009; Fuertes et al., 2012; Deutschmann, Steinvall & Lagerström, 2011). We also adapt our own language to fit underlying norms and preconceived social stereotypes when we communicate with others. In this way, we help to shape individuals through the way we treat them linguistically, and social identity expressed through language is consequently something that is renegotiated during every meeting between humans (Crawford, 1995). An awareness of such mechanisms is especially important for teachers, and other professionals working in the context of human contact.

In most language courses aimed at student teachers of various levels, students are given a theoretical overview of research on aspects related to identity (gender, ethnicity, social class etc.) and language. But however well intended, there is a real danger that research focussed on identifying differences also strengthens stereotypes. Further, there is a risk that such theoretical knowledge remains just that; creating the link between so-called *factual knowledge* – for example, theoretical frameworks and previous studies – and *internalised knowledge*, applicable in our everyday lives, is especially challenging. This is particularly true in the domain of language, where metalinguistic knowledge ideally should be translated into professional language practice, a key skill for anyone working with human interaction.

The following chapter describes experiments conducted in 2011, where we were able to use digital media in order manipulate identity variables such as gender, with the aim to develop and explore experiential pedagogic approaches for raising sociolinguistic language awareness about conceived identity-related phenomena in language. We will also describe the planned second phase of this work, RAVE, where we aim to further develop our models to produce more systematically tested methods for exposing and combatting linguistic stereotyping. We believe that this is an important step in better equipping teachers to judge learners on their individual merits rather than on the preconceived ideas of the group they happen to belong to. Indeed, such awareness is important for any profession related to human contact (health care, the police, law etc.)

2. Previous Research

Stereotyping and Language

Not only does stereotyping, based on various social categories such as gender, age, social class, ethnicity, sexuality or regional affiliation, serve to simplify how people perceive and process information about individuals (Talbot, 2003: 468), it also builds up expectations on how they are supposed to act. People can choose to ignore such expectations, but they still have to relate to them in their interactions with others (Talbot, 2003: 472).

Stereotyping is further complicated by what is often referred to as *intersectionality*. Many researchers argue (e.g. Gutierrez et al., 2012) that aspects of identity such as gender and cannot be analysed in isolation. Negative stereotypes related to different social categories often interact so that the total effect is greater than the sum of individual factors/aspects. In this way, working class black women, for example, may be particularly stereotyped.

Studies have shown that stereotypes and prejudices related to race (e.g. Slaughter-Defoe, 2012) and gender (e.g. Abrams & Rutland, 2008), for example, are established at a very early age and, once learned, they tend to resist change (Killen & Levy, 2008), even when evidence fails to support them or points to the contrary (Sherman et al., 2005).

Language is a key element in this bias. According to Collins and Clement (2012: 377), "language can be conceptualised as a lens that directs and distorts cognition". In spite of social efforts in reducing different forms of prejudice, stereotyping and implicit beliefs remain embedded in language, thereby maintaining hierarchical status relations between groups by distorting people's perceptions in very subtle ways that they may not even be explicitly aware of.

Classic examples of such distortions include gendered implications when generic meaning is intended, leading to interpretations that exclude the other sex (e.g. Bojarska, 2013).

Language is also an important attribute of identity, a signal that draws attention and makes salient certain aspects of the social context. Experiments using so-called matched guise techniques (Lambert et al., 1960), whereby a certain language output is manipulated for regional and/or national accents, have shown that attitudes towards the speaker will be influenced by her/his accent (e.g. Cavallaro & Chin, 2009) and language has in fact been shown to be a stronger stimulus for social categorisation than visual cues such as skin colour (Rakić, Steffens & Mummendey, 2011). Given this close link between language and identity, it becomes an important factor in the definition of boundaries between in- and out-groups.

Stereotyping and language in learning situations

Aspects related to democratic values, identity and communication are clearly defined in all professional degrees of education in the Outcomes section of the Swedish Higher Education Ordinance (Universitets- och högskolerådet, 2011: Annex 2). Such outcomes include the ability of teachers at all levels to: "demonstrate the capacity to respect, communicate and instil a gender equal and equal rights perspective in educational processes"; and to "demonstrate self-knowledge and a capacity for empathy" in this work. Equipping student teachers with such skills presents a real challenge in higher education.

In this pursuit, it is not enough to have explicit knowledge of the mechanisms involved in stereotyping; people's explicit attitudes and intentions, and the influence on their actions and judgments of inbuilt mind-sets, so-called implicit stereotyping, do not always match (e.g. Collins, Biernat & Eidelman, 2009). It is thus especially important for teachers to possess insights into how they themselves are affected by stereotyping structures and how they may inadvertently contribute to these. This motivates a shift of focus from what language differences exist between different social groupings, to what *beliefs* exist about the language behaviour of different social groupings and how these affect us (Edlund, Erson & Milles, 2007).

With specific reference to gender, studies have shown that gendered expectations affect how we experience real events in various learning contexts (Sunderland, 2000). For example, it is well documented that schoolteachers, regardless of gender, tend to give more attention to male than to female students (Sunderland, 2000; Chen & Rao, 2011), even when they think that they are being more attentive to the female students (Sunderland, 2000: 160). A problematised view, however, shows that even if boys get more attention, girls get attention of higher quality, partly due to prejudiced expectations (Sunderland, 2004).

Gender stereotypes also influence students' perception of teachers. Abel and Meltzer (2007) could show that students evaluated a text more positively when they thought that a male teacher had written it and this type of differential evaluation has been replicated in a number of other studies (Centra & Gaubatz, 2000; Godwin & Stevens, 1996). Further, both male and female teachers are more likely to receive better evaluations if they fit gender stereotypes than if they deviate from them (e.g. Basow, 1995; Deutschmann et al. 2011).

Social class and language in learning situations has also been a topic of frequent investigation and debate ever since the proposal of the influential *Code Theory* in the early 1970s (Bernstein, 1971). Although heavily criticised and accused of being "filtered through a strong bias against all forms of working-class behavior" (Labov, 1972), Bernstein's work has inspired many studies that have demonstrated that working class learners are linguistically disadvantaged in education (e.g. Littlejohn, 2002; Atherton, 2002). The language of education is typically based on the middle class sociolect, and individuals do not adhere to the Standard risk being unfairly judged as less intelligent on the basis of their language. According to Littlejohn (2002), this is yet another illustration of how the assumptions of a certain social group are shaped on the basis of their language output. Numerous studies over the past few decades have reported that standard accents are perceived more favourably than nonstandard accents and here non-native and ethnic

accents, in particular, are disfavoured (Edwards, 1999; Lippi Green, 1997; Lindemann, 2003, 2005; Fuertes et al., 2012).

Such tendencies have also been confirmed in educational contexts. Boyd (2003), for example, could demonstrate that non-native speaking teachers in Sweden were ranked low for teacher suitability by a panel of headmasters and pupils on the basis of their accents, although they were highly competent on other linguistic variables such as precision and variation of vocabulary, grammatical correctness and fluency, and had good track records with many years of teaching experience. Accent discrimination has also been widely reported affecting graduate students and instructors at college campuses in the UK and the USA (Bresnahan et al., 2002; Kavas & Kavas, 2008), and similar prejudices seem to be operating on the judgement of learners' performance. Collins et al. (2009), for example, were able to show that teachers translated descriptions of academic performance into lower grades when the student was identified as black. Negative stereotyping has also been noted with regards to LGBTQ students (Crumpacker & Vander Haegen, 1987).

In summary, there is little doubt that linguistic behaviours associated with different social groupings play an important part in how we are judged and in how we perceive others.

3. Case Descriptions

The four cases described below were conducted in 2011 under the project *ASSIS* (A Second Step in Second Life), a project funded by Umeå University, with the aim to use the affordances offered by virtual worlds in order to raise gender awareness among language teacher trainees and also to allow students to discuss gender issues in an international context.

In the project, we worked in *Second Life* (hereafter SL), a 3D virtual world environment, that is built up by its users who communicate through *avatars* (virtual representations of themselves). When students and teachers are represented by their 3D online aliases, unique opportunities for experimenting with identity open up (Warburton, 2009) as participants may choose to be represented by an avatar with, for example, a different race or gender than their usual identification. An added advantage of this method in language teaching is that the anonymity afforded by the environment can reduce anxiety and make quiet students feel more inclined to speak, thereby promoting communication and collaboration (Chester & Gwynne, 1998; Hawisher and Selfe, 1992). There is also the possibility in SL to alter your voice through changing it to a higher (more female) or lower (more male) pitch – so-called voice-morphing. Combining this with the choice of an appropriate avatar produces a convincing illusion, allowing males, for examples, to act through female avatars. In all of the cases described below, we used the voice-morphing tool that is available in SL in packages consisting of five voices (feminine packages, masculine packages etc.) from "Voice Island" – a location in SL searchable in the internal search engine. Once purchased, the voice morph can be activated at the click of a button. You then have the option to activate one of the five voice morphs provided in the package (see Fig. 1 below).[1]

---

[1] For an example of what this may sound like see

Figure 1. Voice morph options in package 'Feminine 2'.

3.1 Matched-Guise Experiment

In our first experiment we explored matched-guise technique using the possibilities that virtual worlds offer in terms of gender manipulation. The original study (see Lambert et al., 1960) investigated how Canadian listeners' attitudes were affected by the language of the speaker. Four bilinguals would read the same text in both English and French. These recordings were then played to respondents or 'judges', who were asked to evaluate the speakers on personal characteristics. Of course the 'judges' did not know that the same people were speaking in the two languages. The study showed that the person reading was evaluated differently depending on which language was used. Since then the technique has been used and developed in a number of studies investigating different attitudes to language output such as national and regional accents, (Giles & Powesland, 1975; Young, 2003; Cavarallo & Chin, 2009). While comparisons of gender effects on evaluations of different dialects have been studied previously (see Andrews, 2003), to the best of our knowledge, no study has explored this technique in virtual worlds and the unique opportunities that they afford the experimenter with regard to gender.

Our experiment was set up in a Master's course at a Swedish university with four female students, from Sweden, Iran, South Africa and China. They were recorded in SL reading a short text using their real voices and female avatars. We then used female-to-male voice morphing and male avatars to record the same students reading the same texts as 'males'. Approximately fifty outside 'judges' were asked to evaluate the avatars using a seven-point Likert scale. Essentially following the methodology and trait inventory of Cavallaro & Chin (2009), the traits were 'hardworking', 'intelligent', 'ambitious', 'confident', 'trustworthy', 'considerate', 'kind', 'honest', 'caring', 'likeable' and 'funny'. One obvious difference compared to previous studies was that the 'judges' could see an avatar. Because the appearance of the avatar could affect the evaluation, the students were asked to keep their two avatars as 'neutral' as possible in relation to one another.

Based on the results from Andrews' study (2003) where male voices were evaluated higher than their female counter-parts on all traits, our hypothesis was that male avatars would be evaluated higher especially on prestige attributes such as 'intelligence', 'confidence' and 'hardworking'. Our hypothesis was refuted. In fact, the female avatars were evaluated higher on **all** characteristics. However, statistically significant differences were only found (t-test, $p \leq 0.05$) for one avatar and for three characteristics ('confident', 'intelligent' and 'kind'). The most feasible explanation for these results was that the poor and artificial quality of the female-to-male voice morphs influenced the 'judges' leading to higher evaluations of the un-morphed female voices. We find it unlikely that the avatar appearances produced this outcome since most of the avatars, male and female, were of rather neutral appearance.

3.2 Students' Gender-Bending

In the second experiment conducted under the project we gave students the opportunity to gender morph to experience if there was any difference in the way that they were treated in conversations

when acting as a different sex. The setting for this second experiment was a Master's course in sociolinguistics where the students were to discuss gender and language matters in a cross-cultural setting with peers from Chile. All participants could choose to gender morph or not, and perhaps because almost all students were female and the female-to-male morphs had proved themselves of poor quality, only one person decided to use this option. Unfortunately, she was far from convincing as the voice sounded artificial, but nevertheless she maintained that the experience was "extremely liberating," but "quite psychologically disturbing," since she was taken aback by how differently she was treated and she became unsure how to respond in this new situation.

In retrospect, the main problems with this model are the ethical dilemmas it presents. Firstly, we do not know how students may react when they enter the role of the opposite sex. Some students may find this extremely disturbing (for a number of reasons), and the experience may trigger psychological processes over which we have little control. A second dilemma is the fact that interlocutors may say or do things they would otherwise not if they knew the true identity of the conversational partner. People may, for example, disclose secrets or make flirtatious approaches, all of which can cause considerable embarrassment and place students in difficult positions. With all this in mind, we hesitate to propose this approach and instead recommend more controlled models where the educators themselves take on the gender morphed roles.

3.3 Evaluation of Teacher Assistant(s)
Study Design
In the third experiment, we used voice morphing in SL to study if gender stereotypes influenced students' perception of teacher performance. The pedagogic aim was to raise teacher trainee students' awareness of this issue. The setting was a course in sociolinguistics on the subject of gender attended by 34 third-year EFL (English as Foreign Language) teacher trainees. The students were to conduct two case discussions on the topic of gender in the classroom in SL, and would be joined by outside 'expert peers' (active language teachers and researchers). Two workshops with group discussions in SL were designed especially for the study. The workshops took place in SL, and each workshop was held twice, once for student group A (discussion groups A1-A5) and once for student group B (discussion groups B1-B5), this in order to create smaller student groups and thereby enabling more teacher-student interaction. Each group was assigned their own "sound parcel" (see Fig. 2). The creation of these sound parcels allowed students to talk privately without being disturbed by noise from other groups.
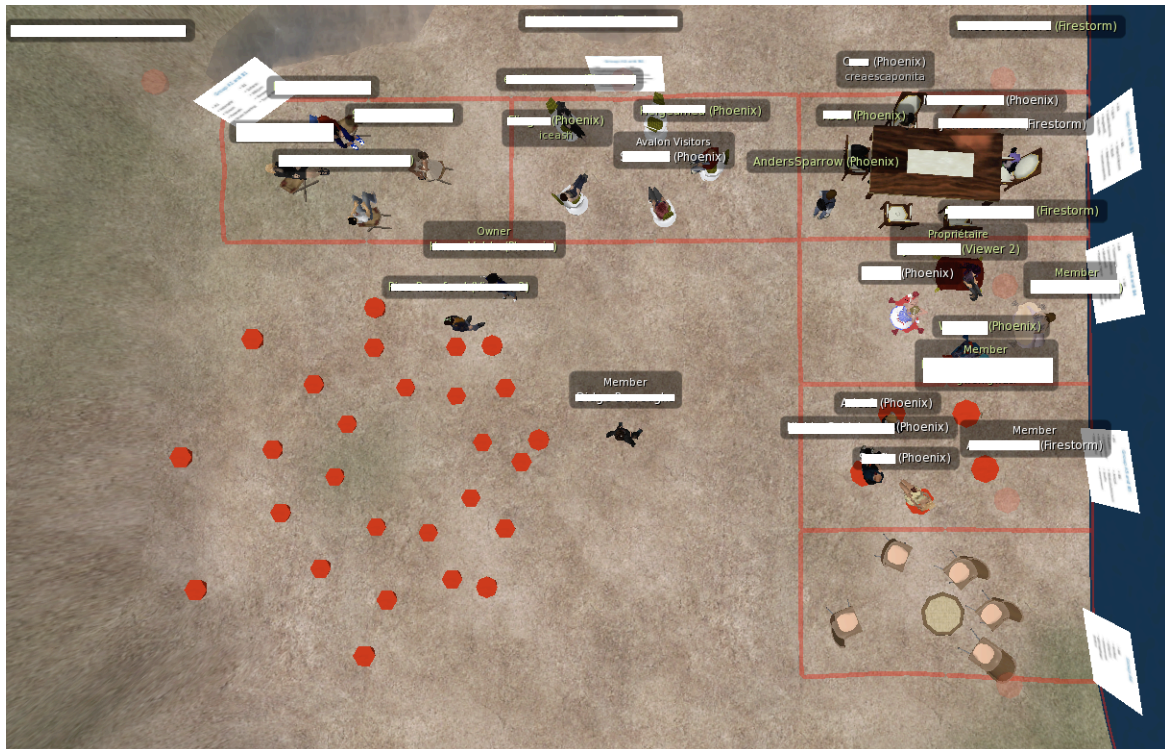
Figure 2. Aerial view of sound parcels used for group discussions. Each red rectangle here represents an area that is sound insulated from the rest of the environment. The students could thus sit in smaller groups and converse without being disturbed by others.

During each workshop various "expert peers" helped to facilitate the students' group discussions as avatars in SL. The "expert peers" were primarily teachers and/or researchers active in language teaching in SL, and these were recruited using online networks such as the SLED-list (Second Life in Education). The rationale behind this was that one of our goals was to give the students an opportunity to discuss issues with educators from different cultural backgrounds.

In order to test the students' prejudices about male and female stereotypes, one teacher (male) used voice morphing to create two "fake" avatars, a woman (Rory) and a man (Rico), who served as teacher assistants during the group discussions in SL. The physical appearance of Rory and Rico in SL were kept neutral but obviously female/male. They both wore jeans and shirts, Rory had long red hair and Rico short brown hair (see Fig. 3). During the study, the students were not aware of the fact that the avatars in fact were the same person but with his voice morphed to a higher (Rory) or lower (Rico) pitch.

Figure 3. The two teacher assistant avatars, Rory (left) and Rico (right).

In the experiment, Rory and Rico took turns walking around in SL to interact with each of the five student discussion groups (groups A1-A5 and B1-B5) during the two hours the workshops lasted for. After the group discussions the students were asked to evaluate Rory and Rico's performances as teachers in an online questionnaire using Surveymonkey (www.surveymonkey.net). Here the students were asked to rate Rory and Rico on a scale from one to six (where six represented total agreement with the statement) in relation to nine statements. The statements included Rory and Rico's performance in facilitating discussion, addressing female vs. male students, and listening vs. talking too much themselves (see Table 1 for full statements used in the questionnaire). At a final meeting the students were also asked to rate how likable and intelligent they thought Rory and Rico were. The student's ratings of Rory and Rico in the questionnaire were analysed using a 2-tailed t-test.

Table 1. The nine statements used in the questionnaire to evaluate Rory and Rico.

1. Overall, Rory/Rico was skilful at her/his job of facilitating the discussion.
2. Rory/Rico addressed male and female avatars equally.
3. Rory/Rico paid more attention to male avatars than female avatars.
4. Rory/Rico paid more attention to female avatars than male avatars.
5. Rory/Rico took over the conversation rather than helping us to engage in discussion.
6. Rory/Rico was good at getting us all to speak without taking too much space herself.
7. Rory/Rico was interested in what I had to say.
8. Rory/Rico had an open mind and listened to all views presented.
9. Rory/Rico had set opinions, which she/he tried to impose on us.

In order to compare the students' ratings of Rory and Rico with actual performance data, the group discussions were recorded in SL (using the software Screenflow), with the informed consent of the students. The recorded material was then used to quantify Rory and Rico's performances by counting the number of times that Rory and Rico respectively addressed female or male students (or expert peers), said something to facilitate the students' discussion or "took over the conversation" by talking about their own opinions or experience. It was also noted whether Rory and Rico spoke to a female or male student or to the discussion group in general. Phrases were considered as "facilitating discussion" if they directly encouraged a specific student or the student group in general to speak. This category included both questions (e.g. "What are your thoughts on that?" or "What statement are you working on?"), and encouraging remarks (e.g. "That's interesting!" or "Go on!"). The recordings were also analysed with reference to instances that Rory and Rico "took over the conversation" and included incidents when Rory or Rico shared examples, experiences or opinions in a relatively lengthy manner that was not directly related to the discussion tasks, even if they were relevant to the discussed topic. Note in this context that the above categories are broad and there was a real risk for interpretation bias. However, since the same person did all the counting for this study the relative personal bias would be similar for both Rory and Rico.

After the two workshops and the evaluations were completed the design was revealed to the students, i.e. that Rory and Rico had been one of their teachers using voice morphing, and the group results from the ratings of Rico and Rory were shown. This information was then used as the starting point for a reflective discussion on attitudes, language and gender.

Results
The results of interest here are how observed conversational performance of Rory and Rico, compared with the students' perception of their performance. In both workshops, both Rory and Rico directed proportionally more questions at female students than male students. This difference was marginal for Rory in Workshop 1, however (see Table 2 below). With regards to facilitating the discussions, Rico produced marginally more facilitators than Rory (39 vs. 38 for Workshop 1

and 42 vs. 33 for Workshop 2). Rory directed a proportionally greater part of her facilitators to female students in Workshop 1, and Rico directed a proportionally greater part of his facilitators to male students in Workshop 1.

Table 2. Observed conversational behaviour Rory/Rico

	Workshop 1 (group A+B)				Workshop 2 (group A+B)			
	female	*male*	*all*	*total*	*female*	*male*	*all*	*total*
No. Participants (%)	28 (67%)	14 (33%)		**42 (100%)**	29 (67%)	14 (33%		**43 (100%)**
Utterances directed at males or females (Rory)	33 (72%)	13 (28%)		**46 (100%)**	21 (84%)	4 (16%)		**25 (100%)**
Utterances directed at males or females (Rico)	34 (81%)	8 (19%)		**42 (100%)**	53 (80%)	13 (20%)		**66 (100%)**
Facilitation directed at males/females/all (Rory)	24 (63%)	4 (11%)	10 (26%)	**38 (100%)**	12 (36%)	4 (12%)	17 (52%)	**33 (100%)**
Facilitation directed at males/females/all (Rico)	14 (36%)	11 (28%)	14 (36%)	**39 (100%)**	18 (43%)	9 (21%)	15 (36%)	**42 (100%)**
Rory "took over" (times)			3				0	
Rico "took over" (times)			6				6	

In Workshop 2, there was no difference in the proportion of facilitators directed at males and females. Note also that several facilitating remarks were directed at the group as a whole. Finally it should be noted that Rico "took over the conversation" on more occasions than did Rory (12 vs. 3 times).

There were several inconsistencies between the observed and student-perceived performance of Rory and Rico (see Table 3 below). When reading the results, note that students were asked to rate performance on a scale from 1 to 6, where 6 meant that they strongly agreed with the statement. For full statements see Table 1. Average ratings for Rory and Rico were compared using a 2-tailed t-test (n.s. = not significant).

Table 3. Student ratings of the performance of Rory and Rico.

	Workshop 1 (groups A+B)				Workshop 2 (groups A+B)			
	Students	Average rating		t-test	Students	Average rating		t-test
Statement	*n*	*Rory*	*Rico*	*p*	*n*	*Rory*	*Rico*	*p*
1. Good facilitator	27	4.48	4.63	ns	23	4.48	4.78	0.016
2. Addressed males and females equally	25	5.28	4.92	0.001	21	5.10	5.10	ns
3. More attention to males	25	1.40	1.72	0.003	21	1.43	1.38	ns
4. More attention to females	25	1.36	1.72	0.009	21	1.57	1.43	ns
5. Took over the conversation	26	2.23	2.12	ns	20	2.30	2.00	0.030
6. Made students speak	26	3.77	4.15	0.001	20	4.25	4.60	0.015
7. Interested in what I	25	4.76	4.40	0.004	21	4.76	4.62	ns

said								
8. Listened with open mind	26	4.65	4.65	ns	20	4.75	5.15	0.002
9. Tried to impose his/her views	25	1.68	1.76	ns	20	1.60	1.80	ns

The students rated Rico significantly higher than Rory for facilitating discussion (workshop 2), giving both males and females more attention (Workshop 1), and for making students speak. This result was unexpected as studies show that female teachers are more likely to be rated high for promoting discussion and giving students attention (Bachen, McLoughlin & Garcia, 1999). However, data from the sound recordings showed that Rico indeed facilitated discussion more, meaning that the students were not influenced by gender stereotypes in this case. Rory, was rated significantly higher than Rico for addressing male and female students equally (Workshop 1), for "taking over conversation" (Workshop 2), and for "listening with interest" (Workshop 1). It is well established that teachers tend to address male students more frequently than female students (Sunderland, 2000). However, the sound recordings showed that contrary to this, and to many students' perceptions, Rory and Rico both addressed proportionally more female students. Similarly, contrary to the evaluations, the sound recordings showed that Rico, not Rory, took over conversation more. It is possible that we expect this kind of behaviour from male but not female teachers and that may have affected perceptions.

The final three statements were not directly linked to the sound recording data and as a result lack the controlling factor or sound recording data. Nevertheless, this data contains the arguably most obvious example of gender stereotyping in this study, i.e. that women are better listeners. That Rory was rated higher for being interested in what the students said (Workshop 1) is consistent with previous studies: Female teachers are often considered to have a better personal connection with the students (Bachen et al., 2009), are perceived to be better listeners (Centra & Gaubatz, 2000) and are considered to give more time and personal attention than their male counterparts (Bennet, 1982). In contrast the students thought that Rico "listened with a more open mind" than Rory did. The literature does not list "open mind" as a male (or female) trait, and we do not know what caused the students to rate Rico higher here. Possibly, to have an open mind is connected with being professional, a trait associated to male teachers according to Spraug & Massoni (2005), but further studies are needed to explore this hypothesis. The statement "Rory/Rico had set opinions which he/she tried to impose on us" was pretty strong and not compliant with the teacher's teaching style, which explains the low ratings for both Rory and Rico in this statement. In summary, the results from the study are complex but at least partly suggest that the students were influenced by gender stereotypes in their evaluation in that Rory was rated higher for being interested in what the students said, for example.

The main purpose of this experiment, however, was not to evaluate differences in perception per se, but rather to expose these and raise awareness of language and stereotyping issues. After the experiment, the design was revealed during a debriefing, with the aim to use the data as a starting point for discussions on gender stereotypes in the classroom and how these had influenced the students' perceptions. Unfortunately, there were too many distractions during the experiment drawing students' attention from the aim, and our students felt that the full intended impact of the experiment as a language raising activity was not realised. There were several reasons for this: Rory and Rico spent very little time with each group since "they" had to interact with all groups; the additional variable of an outside discussion partner took much attention; and finally students were focussing on the content since they had to write two graded reports. Many students claimed that they did not get a lasting impression of Rory and Rico, whose presence they experienced as peripheral. We thus conclude that whilst our results were partly in-line with what we had expected, they did not cause the "aha-effect" we had hoped for. Letting morphed assistants take part during the entire discussion with each group would be a way around this problem.

3.4 The Lecture Model

During the project we were invited by one of our colleagues (Kristy Jauregi at Utrecht University, Holland) to give an online lecture on virtual worlds to her Masters students in Intercultural Communication, studying a course on multilingualism and mediation. In the course, topics such as culture, identity, stereotypes, and the competences of the intercultural speaker and mediator are central. With the aim to demonstrate virtual world identity construction in a practical way, the online lecture, which was formally framed as a talk in SL on virtual worlds by a Swedish lecturer and his two PhD students (see Fig. 3), was projected to two groups of students (36 in all) in a lecture theatre.

Fig. 3. *The male lecturer character (left), with his female PhD student (middle) and male PhD student (right)*

In reality, the lecturer and the PhD students were the same person working from two computers using three avatars, with his un-morphed voice representing the lecturer, a female voice-morph representing the female PhD-student, and a voice-morph which made his voice deeper representing the male PhD-student. During the presentation the lecturer introduced the subject of virtual worlds after which he handed over to "his PhD students" and logged out. In the first trial, they then each gave a 20-minute talk about virtual world projects. After this, the lecturer returned and asked the students to rate the PhD students on the attributes: likeability and intelligence, using a 6-point Likert scale. The ratings were then immediately calculated and the experimental design was revealed. Students were also asked if they suspected that voice-morphing had been going on as, after all, it was partly the topic of the lecture. This was followed by a discussion of the results and what they revealed about the groups' stereotypical views of males and females.

The first group that was given the lecture strongly suspected that some form of voice morphing was taking place (13/20) and hence these results were discarded. In the second lecture we modified the design so that instead of taking two distinct turns, the "PhD students" gave the lecture together in a more conversational fashion, sometimes commenting on each other and asking questions. This was made possible by wearing two head-sets simultaneously and alternatively muting the microphone on one head-set depending on which avatar was supposedly speaking. In this second group no one suspected that the doctoral students were the same person (the lecturer) and there was a significant difference between how the students evaluated the male and the female PhD avatars. The male was evaluated as slightly more intelligent (approaching significance, $p = 0.07$ using a t-test), while the female was deemed as significantly more likeable ($p = 0.001$ using a t-test).

The most favourable results, however, were the responses to the question of what the students had learnt, posed in the post-event survey:

- "I learned how easy it was to influence people's thoughts on somebody's identity/personality […]. It creates a whole new look on how we judge people by looks and gender!"
- "I think it's very interesting how male and female can be so different even though they are the 'same' person."
- "Yes, I learned that even though I think gender isn't important in the vision you have of a person it plays a big role in your valuation of a person, […] in the real world too I guess."
- I think most of us learned that our judgements […] were mostly influenced based on exterior features and voice rather than the information that they gave to us. I think it brought some kind of awareness that you have to look further than only the exterior of a person.

As an exercise in language awareness raising this last experiment was actually the most successful and we have successfully used this model on several occasions after this initial trial. One of the big advantages is that it is relatively efficient and can be conducted during a double lecture. Secondly, since the students do not have to enter the virtual world themselves, it is less technically challenging to 'administer.' Also, as a result of this, we believe that the students can be totally focussed on the actual language event, thereby maximising the language awareness and raising impact.

4. Discussion – Looking Ahead.

The main aim of the case studies described above was to create an 'aha' experience among the students so that they could relate to the theory of language, identity and stereotypical beliefs in a more realistic fashion. As the presentation has shown, this is by no means an easy task. As the project developed, different set-ups were tested and the presentation reflects the chronology of the case studies. Generally speaking, it can be said that the design of the studies developed from the students being the agents and manipulators ('agents'), to us taking those roles and the students instead ending up at the receiving end of the manipulation ('patients'). Table 2 briefly outlines the roles and activities in the four case studies:

Table 2. Roles actions and locations in the four case studies.

CASE STUDY	ARENA FOR MANIPULATION	'AGENT'	ACTION BY 'AGENT'	'PATIENT'	ACTIVITY FOR 'PATIENT'	ARENA FOR 'PATIENT'
Match-Guise Experiments	Second Life	Students	Recording	Anonymous judges	Reacting to recordings	Online
Students' Gender-Bending	Second Life	Students	Interacting with peers	The students' peers	Group discussions	Second Life
Evaluation of Teacher assistants in SL	Second Life	Teacher	Assisting in group discussions	Students	Group discussions	Second Life
The Lecture Model	Second Life	Teacher	Lecturing	Students	Listening to lecture	In lecture hall, watching screen

The development of the case studies was the result of our analyses of each case study with regard to a number of parameters relevant to the educational frame in which they were set and to prior research. Three aspects turned out to be particularly important:

- Time
- Ethical issues

- Quality of data.

From a teaching perspective the time issue cannot be ignored. In our analyses of the case studies, we found that it was far too time consuming setting up environments for students and giving them the necessary guidance for creating and manipulating their avatars and voices. The gains were simply too small in relation to the amount of time invested.

As previously mentioned, the ethical issues in letting students manipulate their identity should not be taken lightly. The effect of the experience of a different identity in a student is difficult to predict and always beyond the control of the teacher. Nor is it possible to control the effects of peer-to-peer interactions when gender issues are destabilised. For these two reasons alone, students' gender manipulation was decided against.

Another aspect of the case studies that turned out to be problematic was the quality of the data. In the two studies where we relied on students' performances, these were simply too inconsistent to generate reliable data. Either the technology itself (voice-morphing) was not reliable, or, as in the second case, the students' use of the technology was rather awkward. Thus, although it appeared to be a good idea to let students subjectively experience stereotypical expectations in different roles, the harsh reality was that this kind of manipulation is difficult to achieve with good quality.

In moving the agency of the manipulation to the teacher and making the students the recipients of the treatment, 'patients' as it were, we hoped to have solved the problems we experienced and to have created a more effective and efficient set-up. Indeed, there are some interesting patterns in the data as shown above. However, when evaluating and analysing the data from the design and the students' answers from the debriefing session, we still found that the environment itself created too much of a distraction in the case where the teaching assistants acted as facilitators in group discussions in SL. As some students were hardly aware of the impact of the assistants, their evaluation of the assistants did not have the desired effect in terms of how they related to the revelation of the manipulation. In short, the debriefing session did not generate the intensity and impact we had hoped for.

In our final case study, we therefore created a scenario in which the students were outside the SL environment and could devote their full attention to the assistants. Such a design guarantees much more control and provides data with little or no interference from other aspects. Accordingly, this case study gave us a model with the control and focus required to generate stable data. Further, the model demonstrated the value of, immediately after the presentations, recording students' impressions and then following them up in a debriefing session. The easily retrieved quantitative data generated qualitative data in the ensuing discussions. Moreover, the students experienced this simple exercise as rewarding and enlightening. To sum up our experiences so far, we can say that although the technology invites all sorts of interesting ideas for awareness raising activities, complex and slightly dirty data may be the result, so there is a good case for keeping it simple; less does indeed seem to be more.

Looking forward, we ask ourselves whether it would not be possible to explore the potential inherent in the digital humanities and virtual experiences in such a fashion that we could raise people's awareness of stereotypes in a more fundamental way. In a new project, *Raising Awareness through Virtual Experiencing* (RAVE), granted by the Wallenberg foundation[2], we aim to do exactly that by building on the experiences from the case studies and creating a model that goes beyond the momentary 'aha' experience. Thus, we hope to achieve a permanent change in our students' awareness of the intricate interaction between social categories, linguistic markers and stereotypical assumptions.

In order to monitor and analyse the long-term effects of awareness-raising activities, a rigorous design is necessary in which independent tests generate data so that we can follow the development of the subjects after the "treatment". This is a new step. Previous studies involving

---

[2] For full details of the grant see: https://www.wallenberg.com/maw/en

match-guise experiments (Giles & Powesland, 1975; Young, 2003; Cavarallo & Chin, 2009, and our own reported above) have had a descriptive approach and recorded people's beliefs linked to the specific event with different representations. It has not been part of the aim to follow the subjects' development. Therefore no calibration has been used that has been linked to an independent test.

The design of the RAVE project includes three independent testing points (pre-test, post-test and delayed post-test), two 'treatments' and two debriefings, of which the second has an extended format and takes place sometime after the 'treatments' in order to allow for higher level and more mature reflections (Watson & Williams, 2004). In recording the debriefing sessions and interviewing participants, the ambition of the project is to assemble qualitative data which will give insights into students' reasoning around theory and reality with regard to stereotypical beliefs linked to language use, and how such reasoning can be changed. Thus, the total data comprises both quantitative and qualitative data which make possible a detailed analysis and comparisons of how awareness-raising activities affect students over time. The ambition is of course that such knowledge, in turn, can help develop procedures for the training of teachers and other professionals working regularly with human contacts so that factual knowledge develops into internalised knowledge; theory could become practice.

References

Abel M.H. & Meltzer A.L. (2007). Student ratings of a male and female professors' lecture on sex discrimination in the workforce. *Sex Roles*, 57: 173–180.

Abrams, D. & Rutland, A. (2008). Gender stereotyping and prejudice in young children: A developmental intergroup perspective. In: Killen, M. & Levy S. R. (Eds.) *Intergroup attitudes and relations in childhood through adulthood*. Oxford University Press, New York.

Andrews, D.R. (2003). Gender effects in a Russian and American matched-guise study: a sociolinguistic comparison. *Russian Linguistics*, 27(3): 287-311.

Atherton, J. (2002). http://www.doceo.co.uk/language_codes.htm

Bachen, C.M., McLoughlin, M.M. & Garcia, S.S. (1999). Assessing the role of gender in college students' evaluations of faculty. *Communication Education*, 48: 193-210.

Basow, S.A. (1995). Student evaluations of college professors: When gender matters. *Journal of Educational Psychology*, 87: 656-665.

Bennett S.K. (1982). Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluation. *Journal of Educational Psychology*, 74: 170-179.

Bernstein, B. (1971). *Class, codes and control, Volume 1*. Routledge and Kegan Paul, London.

Bojarska, K. (2013). Responding to lexical stimuli with gender associations: a cognitive–cultural model. *Journal of Language and Social Psychology,* 32(1): 46-61.

Boyd, S. (2003). Foreign-born teachers in the multilingual classroom in Sweden: the role of attitudes to foreign accent. *International Journal of Bilingual Education and Bilingualism*, 6: 283-295.

Bresnahan, M.J., Ohashi, R., Nebashi, R., Liu, W.Y., & Shearman, S.M. (2002). Attitudinal and affective response toward accented English. *Language and Communication*, 22: 171-185.

Cavallaro, F. & Ng, B.C. (2009). Between Status and Solidarity in Singapore. *World Englishes*, 28(2): 143-159.

Centra, J.A. & Gaubatz, N.B. (2000). Is there gender bias in student evaluations of teaching? *The Journal of Higher Education*, 71: 17-33.

Chen, E.S.L. & Rao, N. (2011). Gender socialization in Chinese kindergartens: teachers' contributions. *Sex Roles* 64: 103-116.

Chester, A. & Gwynne, G. (1998). Online Teaching: Encouraging collaboration through anonymity. *Journal of Computer Mediated Communication*, 4: 2 (online: http://www.ascusc.org/jcmc/vo14/issue2/chester.html).

Collins, E.C., Biernat, M. & Eidelman, S. (2009). Stereotypes in the communication and translation of person impressions. *Journal of Experimental Social Psychology*, 45: 368-374.

Collins, K.A. & Clement, R. (2012). Language and prejudice: Direct and moderated effects. *Journal of Language and Social Psychology*, 31(4): 376 -396.

Crawford, M. (1995). *Talking Difference: On Gender and Language.* Sage Publications Ltd., London.

Crumpacker, L. & Vander Haegen, E. (1987). Pedagogy and prejudice: Strategies for confronting homophobia in the classroom. *Women's Studies Quarterly*, 15(3-4): 65-73.

Deutschmann, M., Steinvall A. & Lagerström, A. (2011).Gender-bending in virtual space: Using voice-morphing in second life to raise sociolinguistic gender awareness. In: Sławomir Czepielewski (Ed.), *Learning a language in virtual worlds: A review of innovation and ICT in language teaching methodology*, International Conference, Warsaw, 17th November 2011. (pp. 54-61). Warsaw: Warsaw Academy of Computer Science, Management and Administration.

Edlund, A., Erson, E. & Milles, K. (2007). *Språk och kön.* Norstedts Akademiska Förlag, Falun.

Edwards, J. (1999). Refining our understanding of language attitudes. *Journal of Language and Social Psychology*, 18: 101-110.

Fuertes, J.N., Gottdiener, W.H., Martin, H., Gilbert, T.C. & Giles, H. (2012). A meta-analysis of the effects of speaker's accents on interpersonal evaluations. *European Journal of Social Psychology,* 42: 120-133.

Giles, H. & Powesland, P.F. (1975). *Speech style and social evaluation*. Academic Press, London and New York.

Goodwin, L.D. & Stevens, E.A. (1993). The influence of gender on university faculty members' perceptions of "good" teaching. *The Journal of Higher Education*, 64: 166-185.

Gutierrez y Muhs, G., Flores Niemann, Y., Gonzalez, C.G. & Harris, A.P. (2012). *Presumed incompetent: the intersections of race and class for women in academia*. Utah State University Press.

Hawisher, G.E. & Selfe, C.L. (1992). Voices in college classrooms: The dynamics of electronic discussion. *Quarterly of the National Writing Project and the Center for the Study of Writing and Literacy*, 14(3): 24-28.

Kavas, A. & Kavas, A. (2008). An exploratory study of undergraduate college students' perceptions and attitudes toward foreign accented faculty. *College Student Journal*, 42: 879-890.

Killen, M. & Levy, S.R. (Eds.) (2008). *Intergroup attitudes and relations in childhood through adulthood*. Oxford University Press, New York.

Labov, W. (1972). Academic Ignorance and Black Intelligence. *Atlantic Monthly* (digital edition). Accessed on 17 September 2013 from: http://www.theatlantic.com/past/docs/issues/95sep/ets/labo.htm

Lambert, W.E., Hodgson, R.C., Gardner, R.C. & Fillenbaum, S. (1960). Evaluational reactions to spoken language. *Journal of Abnormal and Social Psychology*, 60(1): 44-51.

Lindemann, S. (2003). Koreans, Chinese or Indians? Attitudes and ideologies about nonnative English speakers in the United States. *Journal of Sociolinguistics*, 7: 348-364.

Lindemann, S. (2005). Who speaks "broken English"? US undergraduates' perception of nonnative English. *International Journal of Applied Linguistics*, 15: 187-212.

Lippi-Green, R. (1997). *English with accents: language, ideology, and discrimination in the United States*. Routledge, New York.

Littlejohn, S. (2002). *Theories of human communication*. Wadsworth, Albuquerque.

Rakić, T., Steffens, M. C., & Mummendey, A. (2011). Blinded by the accent! The minor role of looks in ethnic categorization. *Journal of Personality and Social Psychology*, 100(1): 16-29.

Sherman, J.W., Stroessner, S.J., Conrey, F.R., & Azam, O.A. (2005). Prejudice and stereotype maintenance processes: attention, attribution, and individuation. *Journal of Personality and Social Psychology*, 89(4): 607-622.

Slaughter-Defoe, D.T. (2012). *Racial stereotyping and child development.* Karger, Basel.

Sprague, J. & Massoni, K. (2005). Student Evaluations and Gendered Expectations: What we can't count can hurt us. *Sex Roles*, 53: 779-793.

Sunderland, J. (2000). New understandings of gender and language classroom research: texts, teacher talk and student talk. *Language Teaching Research*, 4: 149-173.

Sunderland, J. (2004). *Gendered Discourses*. Palgrave Macmillan, London.

Talbot, M. (2003). Gender stereotypes: reproduction and challenge. In: J. Holmes, & M. Meyerhoff (Eds), *The Handbook of Language and Gender* (pp. 468 – 486). Blackwell, Oxford.

Universitets- och högskolerådet (2011). *Higher Education Ordinance – Högskoleförordningen in translation, Annex 2.* Accessed on September 25, 2013 from: http://www.uhr.se/sv/Information-in-English/Laws-and-regulations/The-Higher-Education-Ordinance/Annex-2/

Warburton, S. (2009). Second Life in higher education: Assessing the potential for and the barriers to deploying virtual worlds in learning and teaching. *British Journal of Educational Technology*, 40: 414–426.

Watson, A. and Williams, M. (2004). Post-Lesson Debriefing: Delayed or Immediate? An Investigation of Student Teacher Talk. *Journal of Education for Teaching: International Research and Pedagogy, 30* (2), 85-96.

Young, C.E. (2003). College students' reactions to accents of L2 learners of Spanish and English. In: Sayahi, L. (Ed) *Selected proceedings of the first workshop on Spanish sociolinguistics*, pp. 107-111. Retrieved from http://www.lingref.com/cpp/wss/1/paper1013.pdf