Similarity is More Important than Expertise: Accent Effects in Speech Interfaces

Nils Dahlbäck
Department of Computer and Information Science
Linköping University
SE-581 83 Linköping
Sweden
nilda@ida.liu.se

QianYing Wang
Department of Communication
Stanford University
Stanford, CA, USA
wangqy@stanford.edu

Clifford Nass
Department of Communication
Stanford University
Stanford, CA, USA
nass@stanford.edu

Jenny Alwin
Department of Computer and Information Science
Linköping University
SE-581 83 Linköping
Sweden
jenny.alwin@isv.liu.se

ABSTRACT
In a balanced between-participants experiment (N = 96) American and Swedish participants listened to tourist information on a website about an American or Swedish city presented in English with either an American or Swedish accent and evaluated the speakers' knowledge of the topic, the voice characteristics, and the information characteristics. Users preferred accents similar to their own. Similarity-attraction effects were so powerful that same-accents speakers were viewed as being more knowledgeable than different-accent speakers even when the information would be much better-known by the opposite-accent speaker. Implications for similarity-attraction overwhelming expertise are discussed.

Author Keywords
Speech based systems, Cross-cultural communication, Trust and Liking.

ACM Classification Keywords
H.5.2 User Interfaces, H.1.2 User/Machine systems.

INTRODUCTION
When listening to a spoken message, we not only take in the propositional content of the spoken words. We also automatically and powerfully assess the speaker and classify her on a number of dimensions such as age, gender, and social position [3]. We then use this information to decide how much we will trust the speaker's message [3, 6, 7]. This is true not only for spoken interaction between people, but also true for our interactions with computers using speech interfaces [3].

One key finding concerning voices is the so-called similarity-attraction effect [e.g., 2, 3]. We tend to trust speakers that are similar to us in gender, personality, ethnic background/accent, etc. [3, 5]. For instance, in a study [3, chapter 6] of an on-line e-commerce site, Caucasian-American and Korean participants listened to descriptions of products. Half of them heard product descriptions with Korean accents, and half with American accents. After having heard the descriptions, participants were asked to respond to a questionnaire that asked about the product's likeability and speakers' credibility. The results showed that speakers were rated much more positively when they had an accent that matched participants' ethnicity. And the products they described were rated better in quality and more likable.

Another study [1] had Swedish and American participants interacting with voice interfaces either with an ingroup or an outgroup accent (i.e. English with Swedish or American accent). Their results showed that participants trusted the system with an accent that matched their own more than a system with an accent different from theirs. Users exposed socially undesirable behaviors to a much larger extent and perceived the voice interface more sociable when the accents were matched.

It should, however, be noted, that in these studies above there was nothing *per se* that suggested that a particular speaker was more competent or reliable concerning the information presented or requested (as suggested by the speaker's accent). In this sense, the content was "culturally neutral." But what would happen if an accent different from listeners suggests that the speaker is more knowledgeable about what he or she is talking about? Would one trust a description of some unknown places to see or visit in a foreign city by a local speaker (and hence more knowledgeable) more than a speaker has the same accent as you do (hence more similar to yourself)? In other words, which one will have more influence, expertise or similarity? We designed a study to explore this issue.

EXPERIMENT

The study compared expertise and similarity-attraction using a tourist website to introduce either a U.S. or a Swedish city. The interaction was in English.

Method

The experiment was a full-factorial 2 (participants' country) x 2 (speech output: American accent vs. Swedish accent) x 2 (content: American vs. Swedish tourist information, we used New York City vs. Stockholm) balanced, between-participants design.

Materials

The experimental setting for the study was a fictitious web-site where young travelers post their recommendations about sights to see, places to go, and things to-do for fellow travelers. To control the parallelism of content about New York and Stockholm, we only presented descriptions of non-existent places so that the same descriptions could be used for both cities. This also allowed us to control for participants' prior knowledge of particular places or attractions.

Eight topics of "hidden-treasure" tourist attractions were developed. The eight topics we used were skating, arts and museums, restaurants, nightclubs and pubs, cafés, day trip, swimming/spa, and theatres/musicals. The contents and wordings of these topics were inspired by texts from guide books and tourist information, such as "Everyone has heard of X, but that is a very expensive place to go to. I found, however, that close to there, you can find a …", or "If you decide to visit Y, I must tell you that you can find an interesting but not expensive … ".

For each topic, two descriptions were created, one for New York and one for Stockholm. The only differences between the two descriptions of the same topic were location names, which were tailored to fit either New York (USA) or Stockholm (Sweden). This is one of the texts used.

Restaurants

When you're in Stockholm/New York you've got the chance to experience wonderful food within every category. I prefer Thai food and I will now tell you a little about my favorite restaurant. Take the Red line for Gamla Stan/ Take the 7 train to Queens and stop at Sio Wha. This place really serves sensational Thai food and is so cheap. Order many dishes with all the best from Thai food including seafood, chicken, meat, chilies, curries… and along with that a cool glass of white rice wine. And finish off your meal with Thai pineapple ice cream… and I promise that you will most certainly remember this trip with a content smile.

(Inspired by: Time Out New York, www. nyctourism.com)

Each description was presented by a speaker with either an American or Swedish accent. The American speakers were all U.S. born native English speakers. The Swedish speakers were all native Swedish speakers who learned English as their first foreign language. For most of the Swedish English speakers, their pronunciations were closer to American English speakers than to British or Australian speakers. But they still carry noticeable Swedish accents. From a larger pool of Swedish English speakers, we picked four speakers, two males and two females (same for American speakers), whose accents were noticeably Swedish, but not too strong to understand. For both American and Swedish websites, we assigned different speakers for different topics.

The website was deliberately designed to be not too visually attractive, since we wanted our participants to concentrate on their listening rather than visual aspects of the website.

Participants

Participants were 96 undergraduate students, 48 American and 48 Swedish. All the American students were native English speakers. All the Swedish students were born in Sweden by Swedish speaking parents. They had taken English as their first foreign language in school, which they had studied for 7-9 years, with approximately 2-3 lessons a week. We excluded all participants who had lived for longer periods in an English speaking country.

All Swedish participants were fluent in English for casual conversations, and were used to English textbooks and other English readings. Swedish participants received two movie tickets and American participants received course credit for their participation. Participants were randomly assigned to conditions.

Procedure

Upon arrival, participants read a consent form and signed it. The experimenter presented some initial introduction and gave participants a short tutorial about how to navigate the website. Participants were told that the website was designed for users to submit their vocal opinions for travel suggestions. They were told that they were going to evaluate a website either about Stockholm or about New York, describing restaurants, museums, etc.

The website then guided participants through the evaluation process. After listening to an audio clip of each topic, a window popped up with a few questions about the description they just had heard. They were also asked about their impression of the speaker. All items were based on 1 to 10-point Likert scales involving the general question, "How well do each of these adjectives describe the voice/the system" followed by a series of adjective. The answers were entered via a text questionnaire.

After submitting their answers, participants were led to another description. When they finished all of the eight topics, they were thanked for their participation.

For Swedish participants, the experimental session was followed by a post-study interview. The most important

aspect of this was to make sure that participants met our language requirements for the study. Since asking participants beforehand whether they, for example, had been living in an English-speaking country for over a few months might have alerted participants about the accent aspect of the study, we decided to ask them for their linguistic background afterwards. Twelve participants were excluded because they did not meet our criteria.

RESULTS

All results are based on a full factorial ANOVA. There were no effects for gender of voice or gender of participants.

Familiarity with the city

Participants rated how familiar the speakers were with the introduced city. For New York, participants rated American speakers (M=7.55, SD=1.0) more familiar with the city than Swedish speakers (M=6.91, SD=.1.47), $F(1,46)=3.08$, p=.08. For Stockholm, Swedish speakers (M=7.86, SD=.81) were rated more familiar with the city than American speakers (M=7.14, SD=0.81), $F(1,46)=8.79$, $p<.01$. This measure also served as a manipulation check, demonstrating that participants understood that local speakers were more knowledgeable.

Voice characteristics

We created two indices based on the question "How well do the following adjectives describe the voice?". Both were very reliable. The *voice performance* index consisted of accurate, honest, reliable, trustworthy, competent, creative and intelligent (Cronbach's alpha=0.96). The *voice liking* index consisted of enjoyable, likable, entertaining, helpful, kind, and warm (alpha=0.96).

There was a cross-over interaction effect between participants' origin and speakers' accent, $F(1,88)=6.95$, $p<.01$, with Americans preferring the American voice and Swedish preferring the Swedish voice (see Table 1). Voices from the American speakers were rated higher (M=6.42, SD=1.01) than voices from Swedish speakers (M=6.00, SD=1.18), $F(1,88)=4.2$, $p<.05$, perhaps because all speakers used English. U.S. participants provided higher rankings (M=6.48, SD=1.09) than did Swedish participants (M=5.94, SD=1.08), $F(1,88)=6.72$, $p<.05$.

	Accent	Voice Performance (SD)	Voice Liking (SD)
US Participants	US	6.76 (1.08)	6.97 (1.08)
	Sweden	5.78 (1.38)	5.94 (1.43)
Swedish Participants	US	6.08, (0.84)	5.78 (1.00)
	Sweden	6.21 (0.93)	6.08 (1.06)

Table 1. Results on the voice performance and voice liking measures.

The *voice liking* measure demonstrated a similar pattern to the voice *performance* measure. There was once again a cross-over interaction demonstrating similarity-attraction, $F(1,88)=8.88$, $p<.01$, with Swedish participants preferring the Swedish voice and American participants preferring the American voice. U.S. participants (M=6.37, SD=1.19) gave speakers higher rating than Swedish participants (M=6.01, SD=1.25), $F(1,88)=5.59$, $p<.05$.

Information characteristics

Two indices were created to represent participants' evaluations toward the tourist information provided by speakers. They were based on the question, "How well do the following adjectives describe the information you received?". Both were very reliable. The *information value* index consisted of helpful, reliable, smart, trustworthy, and useful (alpha = 0.95). The *information likable* consisted of entertaining, flexible, friendly, and humorous (alpha= 0.90).

There were cross-over interaction effects—i.e., similarity-attraction effects—between participants' origin and speakers' accent for both the information value factor, $F(1,87)=5.21$, $p<.05$ and the likable factor, $F(1,87)=4.76$, p<.05. American participants rated information from American speakers more valuable and more likable than information from Swedish speakers. Conversely, Swedish participants rated information from Swedish speakers more valuable and more likable than info from American speakers (see Table 2 for more details). There were no main effects.

	Accent	Info Value (SD)	Info Likable (SD)
US Participants	US	6.69 (1.13)	5.97 (1.07)
	Sweden	6.08 (1.30)	5.30 (1.79)
Swedish Participants	US	5.96 (0.94)	5.09 (1.05)
	Sweden	6.35 (0.81)	5.55 (1.02)

Table 2. Results on the information value and information likable measures.

DISCUSSION

The results presented here demonstrate, for the first time, that similarity-attraction effect for voice interfaces are so strong that they overcome other factors that might also affect listener's preferences and judgments, such as expertise.

First, users prefer a voice with the same accent as their own. That is, American participants preferred an American accent and Swedish participants prefer a Swedish accent. Second, participants thought that Swedish speakers knew more about Sweden than American speakers, and vice versa. Thus, not surprisingly, speaker with an accent from

the same county of the introduced city was considered more knowledgeable for that tourist city.

The question, then, is which voice is considered to give more valuable information: the one similar to the listener or the one associated with more knowledge? When it comes to the *information quality* measure, American participants rated information from American speakers higher than that from Swedish speakers. Conversely, Swedish participants rated information from Swedish speakers more valuable than information from American speakers. In other words, the similarity-attraction effect holds for perceived information quality.

There are, at least, two possible causes for this preference pattern. One of course is the similarity-attraction effect, which has previously been shown to influence listeners' preferences for not only speakers' accent suggesting geographic origin, but also, e.g., speaker personality [3]. The other possibility is that when a person matches another's' cultural background, recommendations may be trusted more because there is an assumption of similarity of background. In other words, for a Swedish visitor to New York, even though a person from the US knows more about the city, suggestions from a fellow Swede will be perceived more valuable and more informative. These two possible causes are not mutually exclusive and can work together. Future research should examine cases in which culture would be much less relevant, e.g., teaching contexts.

Regardless of the underlying cause or causes, the preference pattern we found demonstrates that social aspects of communication, in this case, accent similarity-attraction effects, are a very important factor that cannot be ignored when designing voice interfaces for users with various socio-cultural backgrounds. When we think about the computer as a source of information, we must be aware not only of its perceived competence but also the social characteristics that are conveyed through its voice.

ACKNOWLEDGMENTS

The Swedish part of this study was supported by a grant from Vinnova. Magnus Baurén helped us setting up the test site for the Swedish participants.

REFERENCES

1. Dahlbäck, N., Swaamy, S., Nass, C., Arvidsson, F., and Skågeby, J. Spoken Interaction with Computers in a Native or Non-native Language - Same or Different?, *Proceedings of INTERACT 2001,* 294-301 (2001).
2. Byrne, D & Griffitt, W. and Stefanik, D. Attraction and similarity of personality characteristics. *Journal of Personality and Social Psychology*, 5, 82-90 (1967).
3. Nass, C., and Brave, S. *Wired for Speech*. The MIT Press, Cambridge, Mass. (2005).
4. Nass, C. and Lee, K.M. Does computer generated speech manifest personality? An experimental test of similarity-attraction. *Proceedings of CHI 2000*, ACM Press, 229 – 336, (2000).
5. Reeves, B. and Nass, C. *The Media Equation*. New York: Cambridge University Press, (1996).
6. Ryan, E. B. Social psychological mechanisms underlying native speaker evaluation of non-native speaker. *Studies in Second Language Acquisition*, 5 (2), 148-159 (1983).
7. Ryan, E.B., Giles, H., and Sebastian, R.J. An integrative perspective for the study of attitudes toward language variation. In Ryan, E.B. and Giles, H. (Eds.) *Attitudes Toward Language: Social and Applied Contexts*, Arnold, London (1982).