Takayama, L. & Nass, C. (2010). Throwing voices: The psychological impact of the spatial height of projected voices. Proceedings of Computer Supported Collaborative Work: CSCW 2010, Savannah, GA, USA, 91-94.

Throwing Voices: The Psychological Impact of the Spatial Height of Projected Voices

Leila Takayama
Willow Garage
68 Willow Road, Menlo Park, CA, USA 94025
takayama@willowgarage.com

Clifford Nass
CHIMe Lab, Stanford University
450 Serra Mall, Stanford, CA 94305
nass@stanford.edu

ABSTRACT
Communication mediating technologies are throwing our voices away from our bodies in situations ranging from voice conference meetings to mass presentations. Physical height is known to influence dominance in interactions between people [1, 8, 13, 16]. This study explores how audio projection technologies also influence dominance behaviors between people. In an exploratory 2 (between-participants: own voice location set spatially high vs. low) x 2 (within-participants: voice agent set spatially high vs. low) mixed-design experiment (N=64), we investigated the psychological effects of voice location upon collaborative decision-making interactions between people and voice agents. We found evidence that suggests the dominating effects of project voices' coming from above can be mitigated by hearing one's own voice projected from above.

Author Keywords
Throwing voices, vocal height, dominance

General Terms:
Design, Experimentation, Human Factors

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Just as ventriloquists throw their voices into dummies, communication-mediating technologies are throwing our voices away from our bodies in situations ranging from voice conferencing meetings to public announcements and mass presentations in spaces like sports stadiums, concert halls, and churches. Such mediated communication forms are becoming a norm, not only in distant and asynchronous communication settings [3], but also in collocated "live" interactions. Even telephone conversations are described as "the throwing of speech" [7, 17].

Throwing voices describes the decoupling between speech production (i.e., the act of formulating and generating a spoken display) and speech projection (i.e., the act of projecting that spoken display for an audience). The current study investigates the psychological and decision-making implications of decoupling the spatial location of one's voice from one's body via audio projection technologies.

Height and dominance
One important dimension of physical space is vertical height. Social dominance hierarchies typically invoke vertical metaphors of lowness as being more submissive than highness. Based upon the psychological implications of highness for social relations and nonverbal behavior [8], perceptions of power [16], and expressing authority [1], it is plausible that this vertical dimension would influence the psychological experience of thrown voice locations, too.

Previous work on height and dominance has been done in video conferencing and virtual reality settings. One video conferencing study manipulated video monitor spatial locations (high vs. low) between pairs of participants, doing a collaborative decision-making task (Arctic survival); they found that people whose images were presented from the higher monitor were more influential in the decision-making process than those with the low monitor [9]. Another study manipulated height by setting avatar height in virtual worlds (short vs. normal vs. tall) and having participants engage in negotiation tasks; they found that people who had tall avatars split the pool of money more in their own favor; furthermore, people who had tall avatars were less likely to accept unfair offers [18]. Together these studies suggest that perceiving oneself as being short or tall may indeed influence not only others' perceptions of the individual, but also perceptions of oneself and one's subsequent interactions with others.

To study *vocal* spatial height rather than *visual* spatial height, this experiment investigated the effects of spatially high vs. low voices (relative to the listener's head), using an experiment design similar to that of previous work [9].

Power and vocal height
While dominance is the primary construct of interest in this study, it is couched within larger issues of perceived power of voice agents. "[P]ower is broadly defined as the ability to exercise influence by possessing one or more power bases; dominance is but one means of many for expressing power" [2]. Most audio projection systems have speakers placed above the ears of the listening audience, e.g., public announcement systems. Other systems place speakers below the ears of the listening audience, e.g., voice

http://doi.acm.org/10.1145/1718918.1718935

conferencing systems that put speakers on the surface of the conference table. It seems likely that the vertical location of the projected voice will influence its perceived power over those in the physical space, although this has yet to be empirically evaluated.

EXPERIMENT

In this mixed design, 2 (between-participants: own voice spatial location set high vs. low) x 2 (within-participants: voice agent spatial location set high vs. low) experiment (*N*=64), we investigated the psychological effects of voice location upon submissive and dominant behaviors in collaborative decision-making between people and voice agents.

Hypothesis

Because lowness is associated with weakness and submissiveness [13, 16], we predicted that people whose voices were spatially lower than another's voice will be more influenced and feel less influential than people whose voices are of equal or greater height than others' voices.

Participants

Sixty-four university students (32 women and 32 men) volunteered to participate in this study. Their ages ranged from 18 to 29 years, *M*=20.8, *SE*=0.3, and their audio technology experience ranged from none at all (1) to using it several times per week (7), *M*=3.4 (between "several times in the past" and "several times per year"), *SE*=0.2. Their spatial hearing ability scores, as measured by the SSQ [6] ranged from 77 to 175, *M*=125.8, *SE*=3.0. Their trait dominance (14-item index) ratings [12] ranged from -33 to 20, *M*=-6.3, *SE*=1.7. Participants were granted experiment participation credit for their coursework.

Procedure

The core tasks of this study were modified versions of standard collaborative decision-making exercises, the desert and winter survival tasks [10, 11]. Task answers take the form of survival item rankings in order of their importance for survival in the given situation. A participant submits one initial ranking prior to discussion with the voice agent and a final ranking after the discussion.

Prior to coming into the lab, the participant answered a questionnaire about spatial hearing abilities (as used by doctors to pre-screen for patients' spatial hearing impairments [6]), personality (trait dominance [12]), and demographics. Upon arriving in the lab, the participant sat in front of a desktop computer and was given an overview of the experiment session. The study involved testing the audio equipment, doing two collaborative decision-making tasks (one with the spatially high voice agent and one with the spatially low voice agent), filling out questionnaires, and being debriefed on the study. Participants did not have to come to a consensus with the voice agent; the voice agent merely offered one perspective on the situation.

Manipulations

To make the participants' and agents' voices project from low or high spatial locations, computer speakers were placed approximately 45 degrees above and below the ears of the participant. (See Figure 1.)

The agent's voices were projected from either low or high speakers via a computer. Participants were informed that they would be interacting with two voice agents—one with a voice projected from the higher speaker and one from the lower; the speakers were pointed out to the participant during the reading of these instructions.

The participants' voices were amplified and projected from either low or high spatial locations, using the microphone, mixer, and speakers. They practiced speaking into the microphones and listened for their own voice projections before commencing with the study.

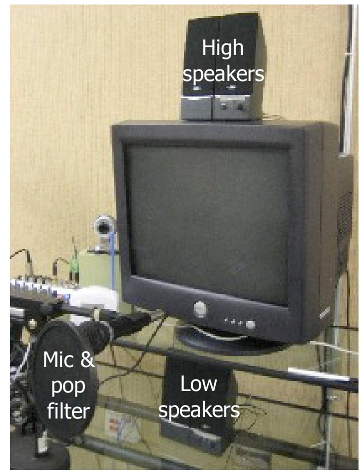

Figure 1. Experiment setting: Participants' ears were brought to the mid-point between speakers; the monitor was left off until it was time to fill out the questionnaire

Experiment design and materials

This study was designed to balance, match, and hold constant all other factors besides the experimental variables of interest. It was balanced for participant gender (female and male), task order (desert survival task first or winter survival task first), and voice agent (pre-recorded voice agent 1 first or agent 2 first). Participant gender was matched with voice agent gender to control for gendered performance effects. Each participant was randomly assigned to an experiment condition.

The voice agents in this study were Wizard of Oz-ed [4]; they were pre-recorded by four voice actors and were played at appropriate times by an experimenter who sat immediately outside of the lab room, listening to the participant via a one-way radio.

Rankings of the items by the voice agent were generated by means of two different algorithms. The algorithms used the participant's initial rankings. For example, in Algorithm A, whatever the participant ranked as #1 was what the voice agent ranked as #5, while in Algorithm B, whatever the participant ranked as #1 was what the voice agent will rank as #3. All voice agents' rankings were at least two counts away from the participant's rankings. Both algorithms provided comparable, moderate degrees of disagreement (*Spearman* ρ=.52, correlation between two ranked lists).

Based upon previous studies [14], the voice agent scripts were designed to be approximately equivalent to each other in terms of utterance length and rhetorical strength. The interaction typically proceeded as follows:

AGENT: Which item did you rank as most important?
PARTICIPANT: The newspapers.
AGENT: I had that item ranked as fifth most important. Why did you decide to rank that as most important?
PARTICIPANT: [explains]
AGENT: The newspapers are rated too high. There is not enough for fire-starting fuel. They will make for poor shelter since they will blow away easily in the desert winds.

Similar exchanges took place for each survival item. If the participant asked a question that the voice agent could not answer, the voice agent did not respond.

Measures of dependent variables

Behaviorally, *how much the voice agent's input influenced the decision of the participant* was measured by the alignment between the voice agent's initial rankings and the participant's final rankings. As in previous work [9], greater alignment was interpreted as less dominant behavior. The distance (sum of squares) between the voice agent's initial rankings and the participant's final rankings measured alignment of the participant to the voice agent's rankings; more distance indicated less submissiveness, i.e., refusing to change one's answers to align with the voice agent.

Attitudinally, *perceived communicative effort* (i.e., how hard people worked to persuade the voice agent) was measured with the Relational Communication Dominance scale [5]. An index was constructed only from items that were interrelated as defined by Principle Components Analysis, including the following five items:

- I attempted to persuade the voice agent.
- I tried to control the interaction.
- I tried to gain the approval of the voice agent.
- I didn't attempt to influence the voice agent. (reversed)
- I didn't try to win the voice agent's favor. (reversed)

These items were highly correlated in both sessions (α=.85 for the first session and α=.82 for the second session), so they were summed to create an index of perceived effort.

Analyses

Repeated measures analysis of variance was used with the voice agent height as a within-participant independent variable, the participant projected voice height as a between-participants independent variable, and experience with audio projection technologies as a covariate. We analyzed the dependent variables of (1) behavioral alignment to the voice agent's rankings and (2) perceived effort exerted to influence the voice agent.

Although trait dominance was measured, it was not a significant predictor of the dependent variables.

RESULTS AND DISCUSSION

Behaviorally, there was a significant interaction effect of own voice location (low vs. high) with agent voice location (low vs. high) upon how much the voice agent's input influenced the decision of the participant, $F(1, 61)=5.98$, $p<.05$. Looking at the simple effects, when the participant's voice location was low, the high voice agent influenced the participant more (M=158.5, SE=21.21) than the low voice agent (M=183.0, SE=21.37), $t(31)=1.97$, $p=.06$ (approaching significance). However, when the participant's voice was high, the agent's voice location was not found to affect the participant's decisions (voice agent low, M=175.6, SE=21.9; voice agent high, M=203.6, SE=23.56), $t(31)=-1.52$, $p=.14$. (See Figure 2.)

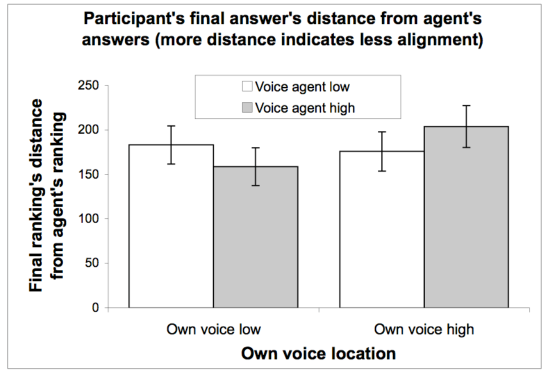

Figure 2. Mean and standard errors for how much voice agent influenced participants' decisions

Attitudinally, there was a significant interaction effect of own voice location with agent voice location on the participant's feelings of trying harder to influence the voice agent, $F(1,43)=4.32$, $p<.05$. Consistent with a low position being less influential, simple effects show that when the person's own voice was low, the participant reported trying harder to influence the high voice agent (M=16.3, SE=1.9) than the low voice agent (M=12.4, SE=1.5), $t(21)=-2.66$, $p<.05$. However, when the participant's own voice was high, then those perceptions were not found to be affected by the voice agent's height, (high: M=14.0, SE=1.4; low: M=14.3, SE=1.39), $t(23)=0.21$, $p=.84$. (See Figure 3.)

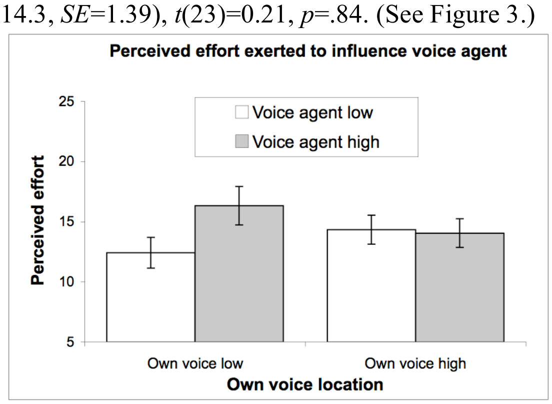

Figure 3. Mean and standard errors for perceived effort exerted to influence the voice agent

These behavioral and attitudinal results support our hypothesis that people whose voices were spatially lower than an agent's voice would be more influenced and feel that they had to try harder to influence the agent than people whose voice agents were at the same height or higher than the agent's voice. However, the opposite was not supported by these data. We had expected to find similar cross-over interactions as in previous work regarding camera angle effects on dominance in video-mediated communication [9]; however, the current results show a weaker trend, which could be because auditory spatial acuity is much weaker than visual spatial acuity in human perception, particularly in the vertical axis [15].

CONCLUSION

In this exploratory study of the psychological effects of the spatial height of projected voices, we have demonstrated how the location of one's projected voice could influence one's own susceptibility to influence. By projecting one's voice from a spatially high location, one can be inoculated against the dominating effects of others' vocal heights.

These results speak to the use of audio projection technologies to empower (being projected from spatially higher locations) or disempower (being talked down upon) individuals when interacting in mediated communication spaces. They suggest that people will behave more submissively to others when they hear their own voices thrown from low spatial locations, but that throwing their voices from high spatial locations may buffer them against the effects of other voice agent heights. Just as video conferencing systems often provide visual feedback about one's own image that is being projected to the distant interlocutors (e.g., a small picture-in-a-picture of one's own video feed), future video conferencing technologies might provide auditory feedback (e.g., sidetones) about one's own vocal image that is being projected to others.

The vertical placement of auditory feedback (i.e., the location of one's projected voice) can affect interpersonal submissiveness in such mediated communication settings. As such, these findings should be considered in the design of interactive systems (e.g., voice-based user interfaces) and mediated communication systems (e.g., voice conferencing systems) when deciding where to place voice projection speakers.

ACKNOWLEDGMENTS

Thanks go to Fred Turner, Jeremy Bailenson, and Herbert Clark for their guidance on this work.

REFERENCES

1. Argyle, M. *Bodily Communication*. Methuen, New York, NY, USA, 1988.
2. Burgoon, J.K. and Dillman, L. Gender, immediacy, and nonverbal communication. in Kalbfleisch, P.J. and Cody, M.J. eds. *Gender, Power, and Communication in human relationships*, Erlbaum, Hillsdale, NJ, (1995), 63-81.
3. Clark, H.H. and Brennan, S.E. Grounding in communication. in Resnick, L.B., Levine, J. and Teasley, S.D. eds. *Perspectives on socially shared cognition*, American Psychological Association, Washington, DC, 1991, 127-149.
4. Dahlback, N., Jonsson, A. and Ahrenberg, L., Wizard of Oz studies: Why and how. In *Proc. CHI 1993,* ACM Press (1993), 193-200.
5. Dillard, J.P., Solomon, D.H. and Palmer, M.T. Structuring the concept of relational communication. *Communication Monographs*, *66* (1999), 49-65.
6. Gatehouse, S. and Noble, W. The speech, spatial and qualities of hearing scale (SSQ). *International Journal of Audiology*, *43*, 2 (2004), 85-99.
7. Gold, R. This is not a pipe. *Communications of the ACM*, (1993), 72.
8. Hall, J.A., Coats, E.J. and BeLeau, L.S. Nonverbal behavior and the vertical dimension of social relations. *Psychological Bulletin*, *131*, 6 (2005), 898-924.
9. Huang, W., Olson, J.S. and Olson, G.M. Camera angle affect dominance in video-mediated communication. In *Proc. CHI 2002*, ACM Press (2002), 716-717.
10. Johnson, D.W. and Johnson, F.P. *Joining together: Group theory and group skills*. Prentice Hall, Englewood Cliffs, NJ, USA, 1991.
11. Lafferty, J.C. and Eady, P.M. *The desert survival problem*. Experimental Learning Methods, Human Synergistics, Plymouth, MI, USA, 1974.
12. Mehrabian, A. Manual for the revised trait dominance-submissiveness scale (TDS).
13. Moeller, S.K., Robinson, M.D. and Zabelina, D.L. Personality dominance and preferential use of the vertical dimension of space. *Psychological Science*, *19*, 4 (2008), 355-361.
14. Moon, Y. Similarity effects in human computer interaction. *Communication*, Stanford University, Stanford, CA, 1996, 149.
15. Rudmann, D.S. and Strybel, T.Z. Auditory spatial facilitation of visual search performance. *Human Factors*, *41*, 1 (1999), 146-160.
16. Schubert, T.W. Your highness: Vertical positions as perceptual symbols of power. *Journal of Personality and Social Psychology*, *89*, 1 (2005), 1-21.
17. Sterne, J. *The Audible Past: Cultural Origins of Sound Reproduction*. Duke University Press, 2003.
18. Yee, N. and Bailenson, J.N. The Proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research*, *33* (2007), 271-290.