Group Decis Negot
DOI 10.1007/s10726-012-9339-x

The Sound of Trust: Voice as a Measurement of Trust During Interactions with Embodied Conversational Agents

Aaron C. Elkins · Douglas C. Derrick

Abstract Trust is a critical component in effective collaboration, decision-making and negotiation. The goal of effective team leaders should be to send signals and messages that increase trust. We attempt to determine if signals can vary perceptions of trustworthiness and if nonverbal behaviors, such as the voice, contain indicators of trust. In order to investigate the relationship between trust and vocal dynamics, this article presents a study that explores how the voice, measured unobtrusively, reflects a person's current level of perceived trust. We used an Embodied Conversational Agent (ECA) to maximize consistency and control in questioning, timing, and interviewer nonverbal behavior, thus eliminating potential confounds that may be introduced due to interaction adaptation. Participants ($N = 88$) completed a face-to-face interview with the ECA and reported their perceptions of the ECA's trustworthiness. The results of the study revealed that vocal pitch was inversely related to perceived trust, but temporally variant; vocal pitch early in the interview reflected trust. The ECA was perceived as more trustworthy when smiling. While the results of this research suggest a relationship between vocal pitch and perceived levels of trust, more work needs to be done to clarify the causal relationship. Similarly, additional study needs to be done in order to integrate additional behavioral measurements that account for variation across diverse situations, people, and cultures.

Keywords Trust · Vocalics · Embodied conversational agent

A. C. Elkins (✉)
Department of Computing, Imperial College London,
Huxley Building, Queens Gate, London SW7 2RH, UK
e-mail: aelkins@imperial.ac.uk

D. C. Derrick
University of Nebraska at Omaha, 1110 S. 67th Street, PKI 172, Omaha, NE 68106, USA
e-mail: dcderrick@unomaha.edu

Published online: 06 January 2013

1 Introduction

Trust is the foundation for functioning relationships, interpersonal interaction, and effective collaboration. Trust between people facilitates the cooperation and mutual adaptation needed to surmount complex and uncertain problems (McAllister 1995; Thompson 1967). A recent study demonstrated the need for establishing trust in negotiation and teams. Tzafrir et al. (2012) found that cooperative negotiators achieved higher joint gains than pro-self negotiators and that this effect was mediated by trust. While teams within an organization may benefit from established working relationships and trust, how can one rapidly integrate new employees or participants into effective collaboration? How can one know if communication is increasing trust?

Consider the following example. During a virtual team meeting with new staff members, the leader would like to encourage reticent employees to be more disclosive and engaged. The manager suspects that participants do not yet trust him or her or the other team members. How should the facilitator try to engage team members and promote trust? Should he or she fill the void by speaking more, be extra friendly, or simply give the participants time to acclimate? Once the manager has selected a strategy, how will he or she know if the strategy is indeed promoting trust during the meeting?

When assessing the affective state of our speaking partners, we attend to their nonverbal and verbal behaviors for vital clues. We usually would consider a smiling and expressive (e.g., using illustrating hand gestures) individual as happier than someone with crossed arms, pursed lips, and a scowling brow. Unfortunately for the manager in our example, the nonverbal behaviors exhibited by people during an interaction reflect more than just their moods and affect. Complex constructs are rolled into the interaction; thus is it difficult for the leader to gauge trust, especially when he or she is also processing and conducting simultaneous leadership activities. However, despite its difficulty, nonverbal and paralinguistic behaviors contain indicators of trust and affect. The challenge is how we can identify them in a timely manner to support managers and facilitators during trust-building periods of collaboration and teamwork.

To investigate the relationship between trust and nonverbal (i.e., paralinguistic) behavior, this study explores how the voice, measured unobtrusively, reflects a person's current level of perceived trust. In order to get to consistent, unaltered interactions from which to measure the voice, we used an Embodied Conversational Agent (ECA) as one of the communication partners. The ECA was included in this study as an interviewer confederate to maximize consistency and control in questioning, timing, and interviewer nonverbal behavior. Specifically, this study explores how manipulating the nonverbal behavior of one communication partner can systematically affect the perceived trust of another person and whether or not these changes in perceived trust can be measured by changes in vocal dynamics. The required degree of consistency in displaying and manipulating nonverbal behaviors could not be afforded without employing an ECA confederate.

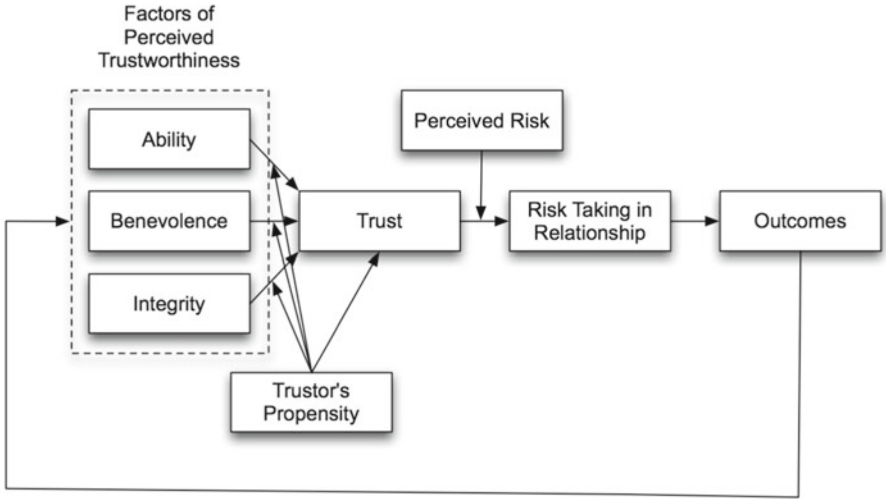

Fig. 1 Mayer's proposed model of trust

2 Theoretical Background

2.1 Perceived Trustworthiness

Research has consistently supported the assertion that trust is multidimensional and consists of many interrelated factors. Some of these include ability (the group of skills, competencies, and characteristics that enable a person to have influence within some specific domain), benevolence (the extent to which a trustee is believed to want to do good to the trustor), and integrity (the trustor's perception that the trustee adheres to a set of principles that the trustor finds acceptable) (Mayer et al. 1995). Figure 1 shows Mayer et al.'s (1995) model of trust. Mayer's model shows how trust is built over time as outcomes feed back into the trustor's perception of trustworthiness.

However, there are very few studies that relate these trustworthiness factors to specific communication strategies or behaviors that can function as signals of trustworthiness (Riegelsberger 2005; Six et al. 2010). Fewer still are studies that try to objectively and automatically measure if these signals of trustworthiness are received and accepted by other people. This is due to the fact that interpersonal communication is complex and as we discuss further, varies from interaction to interaction. One factor that influences the formation of interpersonal trust is the perceived trustworthiness of others (Rusman 2011). Perceived trustworthiness is an individual's assessment of how much and in what context another can be trusted (Hardin 2002). People assess the trustworthiness of another based on information they perceive and/or receive from them. These signs and signals (Bacharach and Gambetta 1997) form the basis of their perceived trustworthiness. Rusman offers the following example of how these signs and signals may work: An academic title such as "professor" (signal) could become a cue for the trustworthiness antecedent "ability", and can lead to the attribution of the trust-warranting property "able" for this specific person (2011). Similarly, a smiling, friendly, and likable demeanor can be a signal or sign of benevolence. Moreover, a recent study showed that due to prescriptive stereotypes, different genders might be perceived as more expert (able), likeable (benevolent), and trustworthy (Nunamaker et al. 2011). A prescriptive stereotype affects perceptions of how a person should be. The theory states that qualities that are ascribed to women and men also tend to be the qualities that are required of them (Prentice and Carranza 2002). Similarly, Ridgeway (2009, 2011) points out how gender frames social interactions and perceptions. For example, there is a prescriptive stereotype that men should be strong and women should be caring (Prentice and Carranza 2002).

This study attempts to consistently manipulate the signals that are sent to a communication partner and to fill the aforementioned research gaps by exploring signals and behaviors that may engender trustworthiness, using the vocal dynamics to predict if these signals were received and accepted. The use of an ECA allows us to vary the gender and demeanor (signals) in a consistent, repeatable way and then measure the differences in perceived trust. We also go a step further to see if automatic and unobtrusive measures can predict the users' perceptions of trust. We use a well-established model of trust to capture overall user perceptions of trustworthiness and track how these perceptions evolve over the course of the interaction. Specifically, we measure the interaction partner's perceptions of ability, benevolence, and integrity and then show how these individual constructs map to the latent construct of perceived trustworthiness. Finally, we show the vocal correlates that are manifested based on this latent construct.

2.2 Interpersonal Communication and Nonverbal Behavior

Interpersonal Adaptation Theory (IAT) (Burgoon et al. 2007) conceptualizes dyadic communication as adaptive and purposeful. When engaging in any type of interaction, people enter with their own set of expectations, beliefs, motivations, requirements, and desires. One of the major predictions of this theory is that people will exhibit a starting nonverbal behavior (speaking voice, demeanor) that may or may not match their partners' behavior. During the course of the interaction, they will seek to minimize the differences in their behavior and synchronize.

In this lens, behavior is not static or universal, but is an adaptation to the behaviors of a speaking partner over the course of the interaction. Because of the dynamic nature of communication, the systematic study of behavior during dyadic interaction can vary from interaction to interaction due to the timing or order of questions, subtle expressions, speaker intonation, and any manner of behavioral influences. In the case of this study, if the speaking partner inadvertently increased speaking volume of his or her voice, participants may increase their volume to match. In this case, the nonverbal change was not a function of underlying affect or emotion, but the tendency towards behavioral synchrony or mimicry (Burgoon et al. 2007; Cappella 1997; Hess et al. 1999; Rapson et al. 1993).

In addition to synchrony, IAT also predicts that time is critical to predicting behavior. Because we are always adapting, our behaviors will be dynamically changing in response to our affect, speaking partner, and environment. To achieve the consistency and time-sensitive control needed for this investigation, an ECA was used as the interviewer confederate. The ECA in this study did not vary its behavior, the order of the questions, or exhibit any inconsistent or additional behaviors that would unsystematically influence participant behavior. Because time is a crucial factor, a repeated-measures experimental design was employed to collect vocal and trust measurements over time and test our first two hypotheses.

H1 Perceived trust in the ECA is temporally variant.

H2 Vocal behavior in response to ECA questions is temporally variant.

These hypotheses test the importance of time and the ongoing interaction of/between behavior and perceived trust.

2.3 Embodied Conversational Agents

The Computer as Social Actors (CASA) theory proposes that human beings interact with computers as though computers were people (Nass et al. 1994). In multiple studies, researchers have found that participants react to interactive computer systems no differently than they react to other people (Nass et al. 1997). It is suggested that people fail to critically assess the computer and its limitations as an interaction partner (Nass and Moon 2000) and as a result, the norms of interaction observed between people are the same as those between a person and a computer (Hall and Henningsen 2008). CASA has been used in multiple studies to provide structure for experimentation. Studies include instances where computers have been specifically designed to praise or criticize performance (Nass and Steuer 1993), display dominant or submissive cues (Moon and Nass 1996; Nass et al. 1995), flatter participants (Fogg and Nass 1997), explore the role of gender and flattery (Lee 2008), or display similar or dissimilar interaction cues with participants (Moon and Nass 1998). Multiple studies have shown that ECA appearance affects users' perceptions. For example, Van Vugt et al. (2008) investigated the effects of facial similarity between users and embodied agents under different experimental conditions. Their results showed that the facial similarity manipulation sometimes affected participants' responses, even though they did not consciously detect the similarity. Niewiadomski and Pelachaud (2010) present an ECA capable of displaying a vast set of facial expressions to communicate its emotional states as well as its social relations. Finally, Qiu and Benbasat (2010) demonstrated that matching ECA and users' ethnicity (not gender) resulted in Product Recommendation Agents being perceived as more social, enjoyable, and useful than mismatched demographics.

Given this prior use of ECAs as a means for social experimentation, it is appropriate that we employed an ECA confederate to function as the communication partner in order to overcome the adaptation effects of IAT. We predicted that the strictly controlled appearance of the ECA would affect the perceived trust of participants and posited our next two hypotheses.

H3 A smiling ECA is perceived to be more trustworthy than a neutral ECA.

H4 An ECA with the same gender as the participant is perceived to be more trustworthy than an ECA with a dissimilar gender.

2.4 Vocalics

Vocalics refer to qualities of speech distinct from the verbal or linguistic content (Juslin and Scherer 2005). Vocalics falls in the category of nonverbal communication referring to "how" something was said instead of "what" was literally said. Linguistics encompasses the literal verbal message, or "what" was said.

Previous research has found that an increase in the fundamental frequency or pitch is related to negative affect, stress, or arousal (Bachorowski and Owren 1995; Juslin and Scherer 2005; Streeter et al. 1977). This phenomenon can be explained intuitively because pitch is a function of the speed of vibration of the vocal folds during speech production (Titze and Martin 1998). For example, females have smaller vocal folds than men, requiring their vocal chords to vibrate faster and leading to their higher perceived pitch. When we are aroused our muscles tense and tighten. When the vocal muscles become tenser they vibrate at a higher frequency, leading to a higher pitch. Previous research has found that when aroused or excited, pitch also exhibits more variation and higher intensities (Juslin and Laukka 2002).

Based on the relationship between negative affect, arousal, and vocal pitch, we predict that when one speaks to someone they do not trust they will speak with an elevated vocal pitch.

H5 Vocal pitch is inversely related to perceived trustworthiness in the ECA.

3 Method

3.1 Overview

Participants in this study completed a 16-question, mock-screening interview conducted by an ECA interviewer. During the course of the interview, participants rated their perceived trust of the interviewer, which served as the dependent measure of this study. The demeanor and gender of the ECA was manipulated during the interview. All of the participant's responses to the ECA's interview questions were audio recorded and used to calculate vocalic measures.

3.2 Sample

Eighty-eight male and female participants from a medium-sized city in the southwestern United States completed the study. The mean age of the participants was 25.45 (SD = 8.44). Fifty-three of the participants were male and thirty-five were female. Eighty-five of them spoke English as their first language.

3.3 Screening Interview Scenario

Replicating the Nunamaker et al. (2011) ECA interview scenario, this study incorporated a screening interview. In this scenario, participants packed a bag of clothes and were instructed to answer the ECA interviewer's questions like they would at

Fig. 2 Embodied conversational agents

an airport. This scenario was employed to facilitate short vocal responses and a one-sided interaction (i.e., ECA asking all of the questions) in a believable and plausible way. This allowed all of the questions and vocal responses to be uniformly asked and answered, with as little variation between participant interviews as possible.

3.4 Embodied Conversational Agent Interviewer

Participants were all interviewed by an ECA. During the 16-question interview the ECA randomly changed its demeanor (smiling or neutral) and gender (male or female). The interview and experiment incorporated a repeated-measures design, as all participants interacted with all four embodied states for four questions during each question block. After every block, participants were prompted to rate their perceived trust of the ECA.

The ECAs used during the experiment are shown in Fig. 2 above. Both the male and female models had similar facial structures and were identical in size, volume, and background. Skin color, eye color, and facial shape are matched between the two models, with the male having a broader chin.

3.5 Procedure

Upon arrival, participants signed and gave informed consent before completing a pre-survey that captured demographic information. Participants were then instructed to pack a bag with various items (e.g., clothes, books) to take through a screening checkpoint. After the participants packed their bag, they approached the ECA station to begin the interview. Once the participant pushed the mouse button in front of the ECA station, the agent asked the first question in the script. The participants spoke their responses and pressed the mouse button when they finished answering the question.

The interview flow can be succinctly described as follows. The ECA first randomly selected a demeanor and gender and asked the first block of questions, after which participants were prompted to rate their trust of the ECA. The agent then chose another demeanor and gender, asked the next question block and repeated the process. The same sixteen questions were asked in the same order every time, but the embodied state (gender X demeanor) was randomly assigned. Each participant interacted with every type of embodied agent. The questions are shown in Table 1 below.

3.6 Measures

The measurement of trust was based on 15 semantic differential word pairs that rate the participant's perceptions of the ECA's integrity, ability, and benevolence. These items have been replicated with high reliability in studies related to source credibility and users' perceptions of ECAs (Ohanian 1990; Reysen 2005) and are shown in Table 2. We collected the user perceptions of the system electronically using semantic differential pairs on a scale from 1 to 7 after each question block.

The first step was to validate the measurements used to evaluate the user's perception of the ECA. Since each participant rated four different ECAs in a within-subjects design, traditional factor analysis, which assumes independence of observations, is inappropriate (Muthén 1991; Rummel 1970). To account for the cluster within participants, the total correlation matrix of the perception measures was partitioned into separate within- and between-subject matrices. The within matrix was then submitted into a multilevel factor analysis using the Maximum Likelihood method and Geomin

Table 1 Questions asked by the embodied agents

1st question block
Please describe in detail the contents of your backpack or purse
I am detecting deception in your responses. Please explain why that is
What will you do after you get through this checkpoint?
Please tell me how you have spent the last two hours before coming to this checkpoint
2nd question block
Has anyone given you a prohibited substance to transport through this checkpoint?
Why should I believe you?
What should happen to a person that unlawfully takes prohibited substances through a checkpoint?
Please describe the last trip or vacation that you took
3rd question block
Do any of the items in the bag not belong to you? If so, please describe which items those are
How do you feel about passing through this checkpoint?
Please elaborate on why you feel that way
Based on your responses, the previous screeners have detected that you are nervous. Please explain why that is
4th question block
Are there any of your responses that you would like to change? If so, please describe what they are
Is there anything that you should have told us but have not?
How do you think that our assessment of your credibility will work out for you today?
Why do you think that?

Table 2 Final dependent measures of users' perceptions

Construct	Semantic differential word pairs	
Integrity	Undependable	Dependable
	Dishonest	Honest
	Unreliable	Reliable
Ability	Unknowledgeable	Knowledgeable
	Unqualified	Qualified
	Unskilled	Skilled
	Uninformed	Informed
	Incompetent	Competent
Benevolence	Unfriendly	Friendly
	Uncheerful	Cheerful
	Unkind	Kind
	Unpleasant	Likable

oblique factor rotation. A factor solution corresponding to ability, integrity, and benevolence were extracted from the within-sample correlation matrix with Eigen values of 6.49, 3.95, and 1.03 ($\chi^2(62) = 119.4$, $p \leq 0.01$, CFI $= 0.981$, RMSEA < 0.01). The CFI and RMSEA statistics suggest a moderately good fit.

Two of the measures of integrity cross-loaded with benevolence and ability and were excluded. All of the ability measures were valid, and four of the benevolence measures were also included. The final measures used for the analysis are shown in Table 2.

The final measures of integrity ($\alpha = 0.87$), ability ($\alpha = 0.94$), and benevolence ($\alpha = 0.95$) were found highly reliable. Given the high reliability of each measure, we computed mean composites for each of the final perception measures.

3.7 Vocal Processing

All of the participants' responses to the ECA's questions were recorded digitally to 48 kHz mono WAV files. The mean length of each vocal response was 7.5 s (SD = 6.15). All of the vocal recordings were resampled to 11.025 kHz and normalized to each recording's peak amplitude. The standard vocal measurement of pitch (F0) was then calculated using the Phonetics software Praat (Boersma 2002).

Because of recording equipment error and poor audio quality, 28 participants had unusable audio. There were a total of 866 audio files processed and included in this study. The final vocalic measurements used in this study were the average vocal pitch and duration across each of the four question blocks for each participant.

3.8 Measurement of Perceived Trustworthiness

After extracting the orthogonal factors of ability, benevolence, and integrity, all of the items listed in Table 2 were submitted to a multilevel confirmatory factor analysis following the protocol suggested by Muthén (1994) and Dyer et al. (2005). Each of

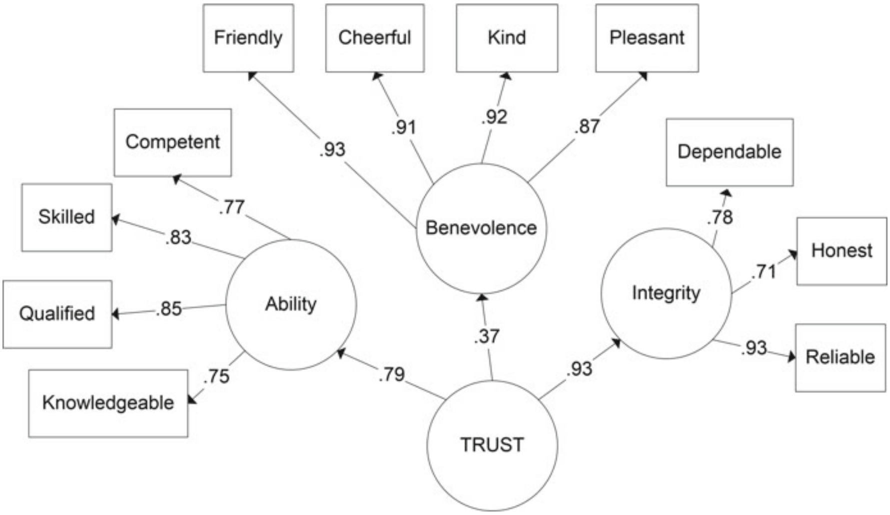

Fig. 3 Confirmatory factor analysis of trust based on within-correlation matrix

these constructs specified in Mayer's Model of Trust (1995) were modeled with paths to a latent variable of Trust.

Both between- and within-subject correlation matrices were simultaneously extracted and submitted to a confirmatory factor analysis using the Maximum Likelihood with full information method (N = 352, Subjects = 88). Intraclass correlations (ICC) measure how much variance in a variable is attributable to between-subject variance (Muthén 1991). ICC for the item measurements ranged from 0.08 to 0.5, suggesting a high degree of between-subject variance that could seriously impact the extraction of factors if subject clustering were ignored. An RMSEA of 0.05 and CFI of 0.974 ($\chi^2(83) = 159.29$, $p \leq 0.001$) indicated that the Mayer measurement model of trust was well fit to these data. The significant χ^2 test likely resulted from the over powered test because of the large sample size (Bollen 1989). Figure 3 illustrates the final measurement Model of Trust and the factor loadings for the within-subject correlation matrix.

Based on the supported factor structure above, an index of trust was calculated by first creating the composites of the factors Ability, Benevolence, and Integrity by averaging. Then, consistent with the supported measurement model, the composites of Ability, Benevolence, and Integrity were averaged to create an index of trust. This measurement of trust is the primary dependent measurement used in the study.

4 Results

4.1 Time and Trust

To assess the relationship between trust and time, a multilevel growth model was specified with trust as the response variable (N = 218) regressed on completion time

Table 3 Comparison of models predicting trust (N = 218, 60 subjects)

	Model 1	Model 2	Model 3	Final model
Fixed effects				
Initial status				
Intercept	4.09***	3.91***	4.02***	3.93***
Rate of change				
Time (s)	0.04**	0.003*	0.005**	0.005***
Vocal pitch * time			9.3e−05*	1.18e−04**
Duration (s)	−0.05***	−0.04***	−0.04***	−0.04***
Avatar smiling		0.48**		0.35**
Avatar male		0.08		
Smiling * male ECA		−0.26		
Vocal pitch (Hz)			−0.01*	−0.01**
Human male			−0.59~	−0.47
No college				−0.93**
Random effects—variance components (standard deviation)				
Level-1:				
Within-subject	0.83	0.80	0.79	0.79
Level-2:				
In initial status	0.54	0.55	0.57	0.50
In rate of change	–	–	–	–
Goodness-of-fit				
Deviance	593.79	583.00	585.58	568.94
AIC	605.79	601.00	603.58	590.93
BIC	626.10	631.47	634.04	628.16

~ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(in seconds) and average question response duration (in seconds) for each question block. To reflect the repeated-measures experimental design over time, both time and the Intercept of trust were modeled to vary within Subject (N = 60) as random effects.

To test hypothesis one that trust is temporally variant and can be predicted by a linear change in time, the specified model was compared to the unconditional means model, which omits any fixed effects using deviance-based hypothesis tests. The difference in deviance statistics was $\chi^2(3, \mathrm{N} = 218) = 19.17$ and significant at the $p < 0.001$ level. This allows us to reject the null hypothesis that time does not predict trust. Allowing random intercepts and time to correlate within subjects did not improve the fit to the data. This means that initial trust levels of participants did not affect the rate of trust change over time.

Examining the coefficients of Model 1 in Table 3 below reveals main effects of time and duration on trust. Participants had an average trust of 4.09 for the ECA at the beginning of the interaction. For every second of interaction with the ECA, trust increased by 0.04, $t(156) = 2.67$, $p < 0.01$, and for every second spent responding over the average of 7.6 s, trust dropped by −0.05, $t(156) = -4.11$, $p < 0.001$.

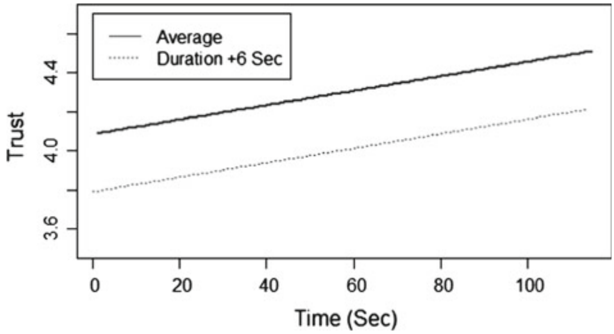

Fig. 4 Main effects of duration and time

Figure 4 illustrates the relationship between time and trust. The average participant over 115 s, increased their trust by 0.42 (4.51–4.09). When participants took 6.1 s over the average response time to respond to the ECA's questions their trust decreased by 0.31, but increased over time at the same rate.

4.2 Time, Demeanor, and Gender

To test hypotheses three and four that the manipulation of ECA Demeanor and Gender affect human trust, dummy coded variables ECA Male (1 = Male, 0 = Female) and ECA Smile (1 = Smile, 0 = Neutral) were added to the growth model. These codes reflect the ECA gender and demeanor participants interacted with prior to reporting their trust levels for each question block.

A deviance hypothesis test comparing the specified model against the growth model reveals a significant improvement to fit, $\chi^2(3, \mathrm{N} = 218) = 10.79$, $p = 0.01$. Providing support for hypothesis three, a significant main effect for smiling was revealed. When the ECA smiled, trust increased by nearly half a point, b $= 0.48$, $t(153) = 2.97$, $p < 0.01$. Discrediting hypothesis four, there was no significant difference between trust of male or female ECAs, $t(153) = 0.53$, $p = 0.59$, nor any interaction between ECA smiling and gender, $t(153) = -1.18$, $p = 0.24$.

Figure 5 below illustrates the effect of Demeanor on trust. While all participants increase their trust of the ECA over time at the same rate, a smiling ECA caused an increase in trust immediately. The figure displays a hypothetical situation of all smiling ECAs versus neutral ECAs. Trust over time could be discontinuous; if Smiling and Neutral demeanors were interspersed throughout the interaction, trust would rise and fall by 0.48.

4.3 Vocal Pitch, Time, and Trust

To test hypothesis five that vocal pitch is inversely related to trust, the vocal pitch of participants while speaking to the ECA was added to the growth model as a fixed effect. The variable Human Male (Male = 1, Female = 0) was included to control

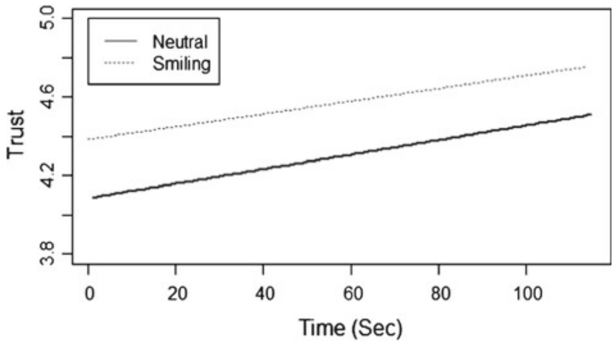

Fig. 5 Main effects of Demeanor and time

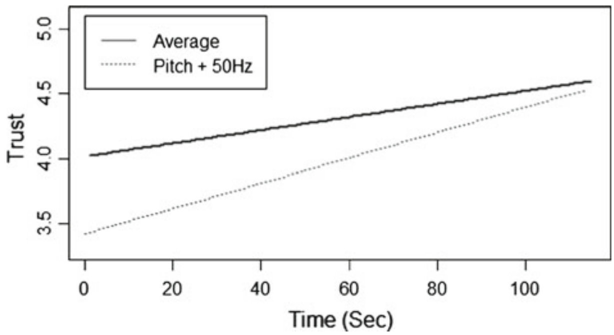

Fig. 6 Main effect and interaction of vocal pitch and time

for the difference in vocal pitch between male and female participants. The deviance hypothesis test revealed a significant improvement of fit to the data, $\chi^2(3, \mathrm{N} = 218) = 8.2$, $p = 0.04$. This allows us to reject the null hypothesis that vocal pitch is unrelated to trust.

Model 3 found below in Table 3 details the relationship between vocal pitch, time, and trust. For every 1 Hz over the average vocal pitch (M = 156 Hz), trust dropped by 0.01, $t(154) = -2.47$, $p = 0.01$. Supporting hypothesis two, a significant interaction of vocal pitch and time was revealed, b = 9.3e−05, $t(154) = 2.19$, $p = 0.03$. This interaction implies that over time the negative relationship between pitch and trust attenuates. Higher vocal pitch earlier in the interaction is more predictive of lower trust levels.

Figure 6 reflects two hypothetical trajectories over a 115 s interaction. The average participant speaking at 156 Hz starts with an initial trust of 4.02 that increases at a rate of 0.005 per second up to approximately 4.6 at the end of the interaction. A participant speaking at 50 Hz above the average pitch (206 Hz) has a lower initial trust level of 3.42, but over time the inverse relationship between vocal pitch and trust attenuates towards equilibrium of trust.

4.4 Final Model of Trust

A final model was specified that includes ECA time, duration, demeanor, vocal pitch, participant gender, and the variable No College (No College $= 1$, At Least Some College $= 0$) to account for some of the participant variance in trust. Using deviance-based hypothesis tests, this model provided a significantly better fit to the data than any of the earlier models and had the lowest model AIC of 568.94. Examining the coefficients in Table 3, we see that the pattern of the predictors remains the same as discussed earlier, however, there was a significant main effect of No College, b $=$ -0.93, $t(57) = -2.58$, $p = 0.01$. Participants that did not have any college trusted the ECA less, but still increased their trust of the ECA at the same rate over time.

5 Discussion

The goal of this study was to manipulate signals that serve as proxies for antecedents of trust and then determine if the changes in users' perceptions of trust could be predicted using the voice. It does appear that trust may have a particular "sound." Initially, the vocalic measures show that the vocal measures of pitch and the duration of response both negatively predicted perceptions of trust. In other words, participants that took longer to respond and answer questions posed by the agent may have done so because of distrust. This distrust may have caused them to feel obligated to explain themselves to the ECA and answer the questions more elaborately. Additionally, vocal pitch was inversely related to trust. However, this effect was strongest earlier in the interaction. Vocal measures of pitch reflect arousal that must be contextualized to interpret. It is quite possible that the participants adapt to the flat and consistent delivery of the ECA over the life of the interaction and there may be some attenuation due to increased familiarity and decreased novelty.

Of all of the individual participant differences, only education level was significant. Age, gender, and other differences were not. However, participants that did not have any college education had a systematically lower level of perceived trustworthiness. This could be based on several factors including their lack of familiarity with technology or the fact they did not view the system as benevolent, or capable. It is conceivable that more educated people had greater confidence in perceived artificial intelligence to posses the antecedent of "ability". This relationship deserves further examination in future studies.

One of the true novelties in this study is using an ECA as a confederate. This allows a degree of fidelity and consistency that would be impossible with a human counterpart. All individuals have biases, attractions, or get tired. Because of IAT, these differences in a communication partner might cause confounds in communication patterns. In other words, we used an ECA because we wanted to be sure that the differences that were detected were solely caused by the change in the manipulation and not some unintended communication difference of the human confederate. The ECA worked well and all participants responded appropriately to the interaction. Using the ECA allowed us to isolate the vocal dynamics and consistently vary the antecedents of trust. The use of ECAs may be a valuable tool in the future to isolate individual effects.

6 Future Research

One limitation of this study is that only the demeanor and gender of the ECA was directly manipulated. While the results of this research suggest a relationship between vocal pitch and perceived levels of trust, more work needs to be done to clarify the causal relationship. A study that specifically manipulates trust should be conducted. It should also be noted that this was a one-time interaction and involved "thin" trust, based on limited interaction. It would be interesting to perform a more longitudinal study to see if participants could build a relationship of trust with the ECA and if a baseline of vocal measurement exists. That is, after repeated interactions, are the vocal measures in a consistent range? If so, it would also be interesting to determine if variances outside of this region are indicative of stress, mistrust, or fatigue.

Further studies that compare these measures to human counterparts are also needed. The goal of this study was to investigate if the voice could be used to predict trust. It would be illuminating to learn whether or not a human could induce similar reactions by varying demeanor and gender in a similar interaction, and if these measures could be accurately and consistently captured. Additionally, further studies with ECAs need to be conducted to ensure that the effects demonstrated here are replicable and to test the ECA's ability to be used in other social science experiments. We posit that the ECA is a reliable confederate, but this needs to be evaluated in future studies.

While the voice shows great potential to provide unobtrusive measurements and accounts for variation in perceived trust, it has its limitations for real-world prediction and classification. The voice reflects predominately arousal, stress, and cognitive load. However, as evidenced by the change in interpretation of vocal pitch later in the interaction, this is insufficient alone for reliable classification of emotions and trust. For example, a person can have an increased vocal pitch, indicating arousal, when both angry and excited. The distinction between the two is a positive or negative valence on top of the degree of arousal. One way to gain more insight into the affect and potential valence would be to analyze the linguistic content of the message spoken. A happier and excited person might use more positive language than an angry person.

When investigating nonverbal indicators of emotion, not all people express behaviors identically, particularly behaviors for which they have less expertise or are practiced at controlling. Someone who is a practiced speaker may actually be facilitated by increased arousal, and their voice may be much less revealing of their emotion than an unpracticed speaker. In order to develop a robust system for real-time classification of trust and emotion, multiple sensors and behavioral features (e.g., vocalics, facial features, language, body gestures) will need to be combined to account for the maximum amount of variation across situations, people, and cultures.

7 Conclusion

One of the major challenges facing collaborative teams or facilitators is reliably monitoring or measuring human cognitive processes. This study demonstrates a model that could 1 day contribute to collaboration software capable of monitoring participant trust, as a function of the voice, time, demographics, and interaction partner demeanor.

The importance of time was revealed. Most importantly, vocal pitch, a measurement of arousal, is dynamic. Any model using the voice to perceive human cognitive states must consider the time and context during measurement. Moreover, within a similar interaction one can expect trust to increase over time, which would inform the development of questions. Questions or discussions requiring greater trust should be placed later in the interaction.

While the interaction context investigated mirrored a screening typical of an airport, the resultant model of trust is generalizable to other interactions that require trust-building phases. For example, collaborations with new team members or a facilitator would have a similar dynamic, requiring the establishment of a first impression and initial trust building. The usage of an ECA provided a reliable interaction that was identical for all participants. Had a human interviewer been used, all participants may have been influenced by the nonverbal behavior or attractiveness, for example. Moreover, the time measurement would have been less reliable because human interviewers ask questions at varying rates and often encourage responses in varying ways through follow-up questions, backchannel nonverbal behaviors, and facial expressions.

Future research should explore further integration of additional sensors and expanded vocal analysis to measure cognitive processing for advanced collaboration software. Work in this area should also cover types of interventions and how to signal team leaders or members when reduced trust or negative cognitive processes are identified. Finally, the ECA included in this study shows great potential for use as an experimental confederate in the social sciences.

References

Bacharach M, Gambetta D (1997) Trust in signs. In: Cook KS (ed) Trust in society. Russell Sage Foundation, New York, pp 148–184

Bachorowski JA, Owren MJ (1995) Vocal expression of emotion: acoustic properties of speech are associated with emotional intensity and context. Psychol Sci 6(4):219–224

Boersma P (2002) Praat, a system for doing phonetics by computer. Glot Int 5(9/10):341–345

Bollen K (1989) Structural equations with latent variables. Wiley-Interscience

Burgoon JK, Stern LA, Dillman L (2007) Interpersonal adaptation: dyadic interaction patterns. Cambridge University Press, Cambridge

Cappella JN (1997) Behavioral and judged coordination in adult informal social interactions: vocal and kinesic indicators. J Pers Soc Psychol 72(1):119–131. doi:10.1037/0022-3514.72.1.119

Dyer NG, Hanges PJ, Hall RJ (2005) Applying multilevel confirmatory factor analysis techniques to the study of leadership. Leadersh Q 16(1):149–167

Fogg BJ, Nass C (1997) Silicon sycophants: the effects of computers that flatter. Int J Hum-Comput Stud 46(5):551–561

Hall B, Henningsen DD (2008) Social facilitation and human-computer interaction. Comput Hum Behav 24(6):2965–2971. doi:10.1016/j.chb.2008.05.003

Hardin R (2002) Trust and trustworthiness. Russell Sage Foundation, New York

Hess U, Philippot P, Blairy S (1999) Mimicry: facts and fiction. In: Philippot P, Feldman RS, Coats EJ (eds) The social context of nonverbal behavior, studies in emotion and social interaction. Editions de la Maison des Sciences de l'Homme, Paris, France, pp 13–241

Juslin PN, Laukka P (2003) Communication of emotions in vocal expression and music performance: different channels, same code? Psychol Bull 129(5):770–814

Juslin PN, Scherer KR (2005) Vocal expression of affect. The new handbook of methods in nonverbal behavior research, pp 65–135

Lee E (2008) Flattery may get computers somewhere, sometimes: the moderating role of output modality, computer gender, and user gender. Int J Human-Comput Stud 66(11):789–800

Mayer RC, Davis JH, Schoorman FD (1995) An integrative model of organizational trust. Acad Manag Rev 20(3):709–734

McAllister DJ (1995) Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. Acad Manag J 38(1):24–59. doi:10.2307/256727

Moon Y, Nass C (1996) How "real" are computer personalities? Commun Res 23(6):651–674

Moon Y, Nass C (1998) Are computers scapegoats? attributions of responsibility in human-computer interaction. Int J Hum-Comput Stud 49(1):79–94

Muthén BO (1991) Multilevel factor analysis of class and student achievement components. J Educ Meas 28(4):338–354

Muthén BO (1994) Multilevel covariance structure analysis. Sociol Methods Res 22(3):376

Nass C, Moon Y (2000) Machines and mindlessness: social responses to computers. J Soc Issues 56(1):81–103

Nass C, Steuer J (1993) Voices, boxes, and sources of messages. Hum Commun Res 19(4):504–527

Nass C, Steuer J, Tauber ER (1994) Computers are social actors. In: Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence. ACM, Boston, MA, pp 72–78

Nass C, Moon Y, Fogg BJ, Reeves B, Dryer C (1995) Can computer personalities be human personalities? In: Conference companion on Human factors in computing systems. ACM, Denver, CO, pp 228–229

Nass C, Moon Y, Morkes J, Kim E, Fogg BJ (1997) Computers are social actors: a review of current research. In: Human values and the design of computer technology, center for the study of language and information lecture notes. Cambridge University Press, Stanford, CA, pp 137–161

Niewiadomski R, Pelachaud C (2010) Affect expression in ECAs: application to politeness displays. Int J Hum-Comput Stud 68(11):851–871

Nunamaker Jr JF, Derrick DC, Elkins AC, Burgoon JK, Patton MW (2011) Embodied Conversational Agent (ECA) based kiosk for automated interviewing. J Manag Inf Syst 28(1):17–49

Ohanian R (1990) Construction and validation of a scale to measure celebrity Endorsers' perceived expertise, trustworthiness, and attractiveness. J Advert 19(3):39–52

Prentice DA, Carranza E (2002) What women and men should be, shouldn't be, are allowed to be, and don't have to be: the contents of prescriptive gender stereotypes. Psychol Women Q 26(4):269–281

Qiu L, Benbasat I (2010) A study of demographic embodiments of product recommendation agents in electronic commerce. Int J Hum-Comput Stud 68(10):669–688

Rapson RL, Hatfield E, Cacioppo JT (1993) Emotional contagion. Cambridge University Press., Cambridge

Reysen S (2005) Construction of new scale: the Reysen likability scale. Soc Behav Pers Int J 33(2):201–208

Ridgeway C (2009) Cultural beliefs and the gendering of social relations. In: Framed by gender: how gender inequality persists in the modern world. Oxford University Press, New York, NY, p 248

Ridgeway CL (2011) Framed by gender: how gender inequality persists in the modern world. Oxford University Press, New York, NY

Riegelsberger J (2005) Trust in mediated interactions (Doctoral dissertation). University College London, London

Rummel RJ (1970) Applied factor analysis. Northwestern University Press, Evanston

Rusman E (2011) How to inform trustworthiness assessments in virtual project teams (Doctoral dissertation). Open University in the Netherlands (CELSTEC), Heerlen, The Netherlands

Six F, Nooteboom B, Hoogendoorn A (2010) Actions that build interpersonal trust: a relational signaling perspective. Rev Soc Econ 68(3):285–315

Streeter LA, Krauss RM, Geller V, Olson C, Apple W (1977) Pitch changes during attempted deception. J Pers Soc Psychol 35(5):345–350

Thompson JD (1967) Organizations in action: social science bases of administrative theory. Transaction Publishers, New Brunswick

Titze IR, Martin DW (1998) Principles of voice production. Acoust Soc Am J 104:1148

Tzafrir S, Sanchez R, Tirosh-Unger K (2012) Social motives and trust: implications for joint gains in negotiations. Group Decis Negot 21(6):839–862

Van Vugt HCV, Bailenson JN, Hoorn JF, Konijn EA (2008) Effects of facial similarity on user responses to embodied agents. ACM Trans Comput Hum Interact 17(2):1–27