International Journal of Innovation in Enterprise System, Volume 3, Issue 02, July 2019, pp 55-63

Towards Modelling Trust in Voice at Zero Acquaintance

Deborah Ooi Yee Hui[1], *Syaheerah Lebai Lutfi[2], Ahmad Sufril Azlan Mohamed[3], and Zahid Akhtar[4]

[1,2,3]Universiti Sains Malaysia, Malaysia
[4]University of Memphis, USA

deborahooi@student.usm.my[1], syaheerah@usm.my[2], sufril@usm.my[3], zahid.eltc@gmail.com[4]

Corresponding author: syaheerah@usm.my
Received: 13 September 2018, Accepted: 16 January 2019, Published: 13 July 2019

***Abstract*—**Trust is essential in many human relationships, especially where there is an element of inter-dependency. However, humans tend to make quick judgements about trusting other individuals, even those met at zero acquaintance. Past studies have shown the significance of voice in perceived trustworthiness, but research associating trustworthiness and different vocal features such as speech rate and fundamental frequency (f0) has yet to yield consistent results. Therefore, this paper proposes a method to investigate 1) the association between trustworthiness and different vocal features, 2) the vocal characteristics that Malaysian ethnic groups base their judgement of trustworthiness on and 3) building a neural network model that predicts the degree of trustworthiness in a human voice. In the method proposed, a reliable set of audio clips will be obtained and analyzed with SoundGen to determine the acoustical characteristics. Then the audio clips will be distributed to a large group of untrained respondents to rate their degree of trust in the speakers of each audio clip. The participants will be able to choose from 30 sets of audio clips which will consist of 6 audio clips each. The acoustic characteristics will be analyzed and com-pared with the ratings to determine if there are any correlations between the acoustic characteristic and the trustworthiness ratings. After that, a neural network model will be built based on the collected data. The neural network model will be able to predict the trustworthiness of a person's voice.

***Keywords*—**prosody, trust, voice, vocal cues, zero acquaintance.

I. INTRODUCTION

According to the Cambridge English Dictionary, to trust someone is "to believe that someone is good and honest and will not harm you" [1]. Trust is central to various commercial and interpersonal relationships. In fact, trust is essential wherever risk, uncertainty, or interdependence exist. Without the presence of trust, negotiating deals would be much more difficult as people would lack trust in the other party. Similarly, a student's trust in a teacher and vice-versa is a crucial factor in learning and teaching process [2]. Namely, trust is vital in enabling maximal learning in a teacher-student relationship [2], [7].

However, trustworthiness can also actually be gauged in a zero-acquaintance situation [8], [9]. Zero acquaintance is a situation in which perceivers make judgments about targets they are given no opportunity to interact with [10]. [11] mentioned that intuitive judgments at zero acquaintance are strong predictors of the performance of teachers [12], telephone operators [13], salespeople [8], and other professionals [14].

In recent years, several researches have been conducted on the factors that determine trustworthiness. Particularly, many studies have shown the significance and importance of voice in perceived trustworthiness. For instance, from a medical perspective, a research has found that verbal exchanges between neonatal-care nurse and mother greatly influences a woman's confidence, her feelings of connection to her infant and her sense of control [15]. Another research conducted by [16] supports that fact by indicating that audio communication shows a more significant emergence of trust as compared to text chat. Whereas, [9] studied facial features and identified that men with greater facial width were more likely to exploit the trust of others and that people were more likely to trust male counterparts with narrow rather than wide faces [9].

Despite recent few researches that analyzed the effect of various vocal cues (e.g., phonetic segmentation, prosody, pitch, speech rates, and fundamental frequency (f0)) on trustworthiness, research addressing the factors of trustworthiness at zero acquaintance is scarce and limited. The state-of-the-art in voice characteristics on trustworthiness at zero acquaintance is nascent. Especially, we are still far from consensus regarding the direction of association between different vocal features - in particular, speech rate and fundamental frequency (f0) - and trustworthiness. On one hand, a deception study indicated that participants consistently raised their f0 when lying [17]. While another study indicated the opposite [18]. Yet another further study found no differences in f0 or other acoustic characteristics when analyzing deceptive and truthful messages [19]. This situation highlights the fact that there is not yet any consistent results from research regarding the direction of association between different vocal features - in particular, speech rate and fundamental frequency (f0) - and trustworthiness. Thus, there is a need to investigate deeper into this area. This paper plans to suggest the method to investigate the association

between vocal features and trustworthiness at zero acquaintance. Besides that, the method suggested in this research will also aim to identify the characteristics that different ethnics in Malaysia base their judgement of trustworthiness on. Furthermore, this paper suggests a method to predict the degree of trustworthiness in the voice of an individual.

It is important to note that this paper documents the research while it is still at the proposal stage. This research has yet to be carried out fully, and therefore is unable to deliver any conclusive results yet. However, this paper does suggest an approach that could be carried out in order to investigate the effect of vocal cues on trustworthiness. Throughout the rest of the paper, the term "research/study/investigation" will refer to the overall re-search topic and objective (to investigate the effect of vocal cues on trustworthiness) and the term "paper" will be used to describe this research paper.

The paper is organized as follows. There is a total of 6 sections. In Section 1.1, the problem statement is discussed. The objectives of this research are clearly stated in Section 1.2 and the benefits and impact of the study is elaborated in Section 1.3. Section 2 contains the literature review of related work that has been conducted in this field of re-search. Section 3 discusses in detail about the project methodology. The expected deliverables are clearly stated and elaborated in Section 4. Section 5 contains the conclusion of this paper and Section 6 states the acknowledgements of this research. Section 7 contains the references in APA style citation.

1.1.Problem Statement

Although there has been quite a few research works that analyzed the effect of various vocal cues on trustworthiness, research addressing the factors of trustworthiness at zero acquaintance are few and far in between. Research indicating the impact of voice characteristics on trustworthiness at zero acquaintance has yet to be fully developed. Essentially, we have yet to reach a unanimous result with regards to the relationship between different vocal features - in particular, fundamental frequency (f0) and speech rate - and trust-worthiness. Thus, there is a need to investigate deeper into this area. In order to conduct further investigation into this area, a suitable method must first be obtained.

1.2.Objectives

The objectives of this research are as follows:

1. To investigate the association between vocal features and trustworthiness at zero acquaintance.
2. To identify the characteristics that different ethnics in Malaysia base their judgement of trustworthiness on.

1.3.Benefits / Impact / Significance of Study

There are many impacts of conducting this study but ultimately the main aim of this study is to facilitate online tutoring.

In this present age, online tutoring has become more and more common and many tutors are turning to online tutoring for extra income. However, there exists a barrier between the students and the teachers, where the student is not completely sure that he or she can trust the teacher, and likewise the teacher to the student. In fact, this research is part of an ongoing project to produce a trustworthy robotic online tutor. Research is currently underway for detecting the impact of different facial features, gestures and even handwriting on trust. This research that investigates the impact of vocal features on trust at zero acquaintance will ultimately be used when producing the voice of the robot. Currently, most robots are programmed to use synthetic speech as their voice but inherently there exists a lack of trust when it comes to synthetic speech. Therefore, this study will hopefully be able to contribute towards making a more trustworthy robotic voice that will maximize the teaching and learning capability in online tutoring.

II. RELATED WORK

Trust is central to various interpersonal relationships and essential wherever uncertainty, risk, or interdependence exist, such as in commercial relationships. However, there cannot be trust without integrity [20]. According to [21], integrity is the basis of trust, and it varies according to the individual. A number of research has emphasized the importance of voice in perceived trustworthiness, such as [22], [23], [16]. Various studied have analyzed specific features of voice that contribute to an overall increase in perceived trustworthiness. A brief tabularized overview of all the related work is shown in Table 1.

TABLE 1.
A BRIEF OVERVIEW OF ALL RELATED WORK

Author	Area of Research	Features	Outcome
[22]	Consumer Trust	Effects of Voice and 3D Avatars on Consumer Trust	Presence of voice increased trust
[23]	Impact of Media on Consumer Experience	Effects of different media such as telephone, television, web and trust on consumer experience	Using telephone increased trust
[16]	Trust in Four Different Communication Situations	Face-to-face, video, audio, text chat, trust	Ratings of trust in face-to-face > video and audio > text chat

[24]	Trust in the Workplace	Trusting stance, Organisational Membership Trust, Hierarchical Relationship, Gender	Trusting stance was positively related to initial trust level.
[25]	Trust within Health Care Provision Relationships	Key dimensions of trust in health care provision relationships (patient and provider, health worker and employer) using South African evidence	Respectful treatment is the central demand of primary care service users
[26]	Trust in Market Research Relationships	Factors that determine users' trust in their researchers, such as individual, interpersonal, organizational, interdepartmental, and project factors.	Interpersonal factors are the most predictive of trust.
[20]	Research Integrity, Trust in Published Research	Constraints or Influences on Researcher Behaviours, Research Integrity guidelines	Respect for persons, beneficence, and justice should be primary ethical principles; Peer review is important to ensure integrity.
[27]	Trust in Virtual Systems	A trust model based on word-of mouth mechanism, grounded in real-world social trust characteristics.	The trust model allows agents to gauge their trust and understanding of other agents' opinions.
[28]	Antecedents and Effects of Trust in Virtual Communities	Measurement of ability and benevolence/integrity by applying an existing scale to data collected from virtual communities	Trust effects a virtual community member's desire to get information more than to give information.
[29]	Trust in E-commerce	A multidimensional-trust model for online exchanges in B-to-C ecommerce; relationship between academia's trust perspective and practitioner's trust perspective	The two perspectives are divergent in some ways and complementary in other ways.
[30]	Trust in Consumer Internet Shopping	Influences of various factors on consumer trust in Internet Shopping	Merchant integrity is a major positive determinant of consumer trust in Internet shopping
[9]	Validity of Facial Cues in Determining Trustworthiness	Effect of Male Facial Width on Perceived Trustworthiness	Wider-faced men were more untrustworthy, Narrow-faced men are trusted more.
[36]	Determining Trustworthiness from Voice	Detection of Past Infidelity from Voice	Lower-pitched voices were rated to be from cheaters, human voice may be useful in detecting cheating
[37]	Using Voice Acoustics to Determine Personality Ratings	Using data-driven voice computational modelling to investigate effect of acoustics on perceived trustworthiness	Ratings could be modulated by manipulating acoustical difference; There is a strong acoustical basis for voice personality impressions
[38]	Prosodic Cues and Trustworthiness	An investment-game method to assess impact of prosodic characteristics on trust attributions	Speaker accent, mean pitch, and articulation rate all influence participants' trust
[18]	Facial and Vocal Cues of Deception and Honesty	The extent to which nonverbal cues indicate deception	Honesty/deception is more accurately conveyed in a posed condition than spontaneous.
[39]	Usage of Voice to Convey Attitudes	Possible acoustic cues of sarcasm by native English Speakers	A specific pattern of prosodic cues can reliably indicate sarcasm
[40]	Speech and Personality Perception	Effects of Speech Rate on Personality Perception	Faster voices are considered more competent, but less benovelent. Slower voices are considered less competent and benovelent.
[41]	Effects of Voice on Perception and Personality	Impact of only pitch on perception and personality	Higher-pitched voices are considered more extraverted; it is quite impossible not to consider semantic content
[42]	Effects of Prosody on Personality Features	Effect of pitch level, pitch range, articulation rate and loudness in synthetic speech on branding personalities	Modelling personality in synthetic speech is possible.
[45]	Paralinguistic Cues and Perceived Trustworthiness	Effects of Visual and Paralinguistic Cues on Confidence, and Perceived Trustworthiness	There is a superiority effect of facial over paralinguistic information
[43]	Voice Accent and Perceived Trustworthiness	Voice Accent, Local speakers, Foreign Speakers, Trust	A foreign spokesperson is perceived to be less trustworthy than a local spokesperson.
[44]	Emotions and Perceived Trustworthiness	Emotions, Trust	Trustworthiness is influenced by the emotions displayed.

A pool of research been conducted on factors influencing trustworthiness. However, most prior works are related to trust in the workplace [24], [25], trust in research integrity [26], [20], trust in virtual systems [27], [28], or trust from an e-commerce perspective [29], [30]. Trust in the workplace are mostly based on employee past experiences, current

working conditions, etc., while trust from an e-commerce perspective are based on the transaction's value and perceived risk.

Although there are many aspects of trust, the decision to trust or not to trust still relies on human judgement. [8] has stated that an abundance of information can be determined by social perceivers about other people through only brief exposure to their expressive behavior. [31] supports that statement by proving that observers are able to make accurate judgements about others even in a situation where there is an absence of any personal interaction between the targets and raters. More research on this topic has proven that even impressions formed on very quick observations have been found to be accurate or predictive. This situation is true especially for certain personality traits [32], [10], [33].

Samples from any available channel of communication, such as the body, transcripts, the face, the voice, speech, or any combination of the previously mentioned channels are regarded as "thin slices" and may be shown to research participants. Thin slices have been shown to retain most of the fluid and dynamic information while removing information about the context of the situation, the ongoing verbal stream or any history of targets. Therefore, thin slices remove the influence of the verbal message or information from previous interactions from the observer's perspective and thereby causes the observer to pay attention to the nonverbal cues. Most of the time, only thin slices of behavior are all that is needed for humans to quickly make personality trait judgements [8]. According to [8], a thin slice is defined as any excerpt of dynamic information less than 5 minutes long.

[31] has coined this situation in which perceivers make judgments about targets they are given no opportunity to interact with as "zero acquaintance". Several other researches have been conducted previously with regards to zero acquaintance [11], [14], with each experiment utilizing thin slices that last from 5 seconds to 30 seconds each. In a successful study conducted by [34], only short 10-second audio clips were used to detect the effects of prosodic features of voice on personality. Therefore, the 10s length could be sufficiently long enough to capture personality impressions while lightening the load on the listeners. In addition, it is interesting to note that increased attention reduces the accuracy of zero acquaintance judgements. In other words, accurate thin slice judgements are disrupted by deliberate and conscious processing [14].

Although there has been much research previously conducted on determining personality traits at zero acquaintance, research addressing the factors of trustworthiness at zero acquaintance is still scarce and limited. Nonetheless, there are a few studies that claim that factors such as facial features and voice affect the perceived trustworthiness of an individual at zero acquaintance. For example, a research conducted by [9] identified that men with greater facial width were more likely to exploit the trust of others and that people were more likely to trust male counterparts with narrow rather than wide faces. This is not considering the male counterpart's attractiveness.

With regards to the effect of voice on perceived trustworthiness at zero acquaintance, a pool of research has already been previously conducted. [35] has indicated that user affect could be modelled solely by using conversational features. This conclusion has been empirically proven through their investigation, which uses conversational features to predict satisfaction ratings in Human-Computer Interaction. One such example of user affect could potentially be trust. [36] has observed the likelihood that an individual has cheated on committed, romantic partners simply by hearing 10 seconds the speaker's voice. In addition to comparing the original voice samples of male and female voices against each other, they also manipulated the pitch of the voice samples to identify and link between the difference in pitch and the likelihood of cheating. In their research, they discovered that male speakers were rated as "more likely to have cheated" and that female raters gave higher ratings for cheating than men. They also observed that both males and females with a manipulated lower pitch had higher cheating ratings than those with manipulated higher pitch. As [36] noted, pitch does play a role in detecting infidelity, but it does not represent the entire picture, since raters were still able to identify cheaters by only using very thin slices of vocal information.

[37] on the other hand, utilized data-driven voice computational modelling to study the relationship between acoustics and perceived trustworthiness in the single word "hello". By exaggerating or reducing the acoustical difference between the prototypes, the difference in trustworthiness ratings could be modulated quasilinear, resulting in a strong caricaturing effect. Hence, a strong acoustical basis for perceived voice personality is clearly demonstrated. Personality impressions such as trustworthiness can be accurately modulated parametrically by acoustical changes, and considerably increased (or decreased) by caricaturing.

Research has also been done with regards to using prosodic features to measure trustworthiness. [38] for example, has assessed how trust attributes are affected by the prosodic characteristics of speakers of a range

of British English accents by using the investment game. In their research, they have observed that higher investments were associated with faster articulation rate and higher pitch, which is consistent with some, but not all previous studies on attributions of trustworthiness and associated characteristics [18], [39], [42].

[43] has also indicated that accent plays an influential role in determining a person's trustworthiness. According to [43], a spokesperson with a foreign accent would be rated as less trustworthy than a spokesperson with the standard accent of the audience.

Furthermore, [44] states that perceptions of trustworthiness of an individual are influenced by the emotions displayed by the individual. In terms of para-linguistic cues, [45] researches the relative role of facial and vocal (para-linguistic, non-verbal) information in processing information about trustworthiness. However, her findings concluded that a superiority effect of facial over paralinguistic information exists. In other words, from a paralinguistic point of view, people tend to judge trustworthiness via facial expressions much more than they do with vocal cues.

Although much research has already been conducted on the effect of different vocal cues on perceived trustworthiness of an individual, the role of voice characteristics and prosody, in particular, with respect to attributions of trustworthiness are limited. There are several issues yet to be investigated properly to establish consensus about interrelationships among different vocal covariates, trustworthiness and integrity. This research aims to fill some of these limitations of prior works. Since this paper is still at the proposal stage of the research, this paper aims to describe and suggest a method to conduct the research.

III. METHOD

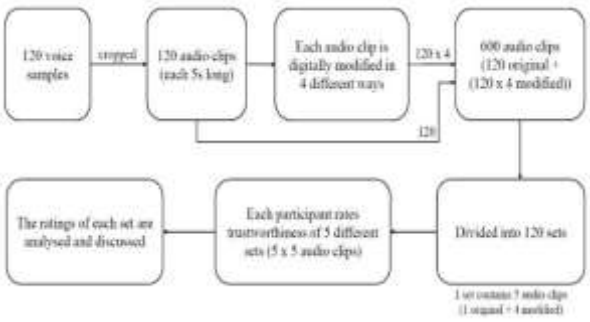

Fig. 1 A general overview of the previous investigation flow

Initially, the original approach to conduct this investigation was as shown in Figure 1. The original approach involved a total of 120 different voice samples from the OMG-Emotion behavior dataset that have been cropped to form a 5 seconds long audio clip. Each audio clip would be subjected to acoustic analysis using Praat speech analysis software and then digitally manipulated in 4 different ways, namely increasing and decreasing the rate and f0 respectively. This will result in a total of 600 audio clips being formed (120 original +(120*4 copies)). These audio clips would be divided into 120 sets respectively. Each set would contain one original audio clip, and 4 digitally manipulated versions of that audio clip.

Each participant in this study would then be asked to rate the trustworthiness of 5 different sets on a scale of 1 to 10 (1 being least trustworthy and 10 being most trustworthy). The participant would submit their ratings via a simple online assessment. Since Peninsular Malaysia consists of 3 main ethnic groups, namely Bumiputera, Indian and Chinese, this research would also include participants from all these races. The ratings given by each ethnic group would be recorded and analyzed accordingly.

Finally, the ratings of each set would be analyzed. There are two ways in which the sets would be analyzed. Firstly, the effects of rate and f0 on the trustworthiness ratings, as compared to the original will be analyzed. This scope of analysis would be termed "intra-set". Next, the effects of the semantic content of the audio clips would also be analyzed. This could be carried out by investigating and comparing the ratings between sets. This scope of analysis would be called "inter-set". The difference in ratings between each ethnic group would also be evaluated and the results would be documented, analyzed and discussed. This approach was pilot tested to check the feasibility and validity of the approach. The pilot test was carried out with 5 audio clips, each digitally manipulated 4 different times. Through trial and error, the margin in which the voices were manipulated significantly yet still sounded realistic was determined to be +-10% of the original f0 and +28%, -14% of the original speaking rate. As this margin was determined from the perspective of the author who is also the pitch manipulator, the degree of accuracy of the margin is questionable and had to be verified. To verify the opinion, these 25 audio clips (5 original + 20 manipulated) were given to 5 independent raters. Meanwhile, the audio clips were acoustically analyzed using Praat and annotated. The 5 independent raters were asked if a) the voice clips sounded realistic and b) to rate their degree of trust in the speaker. Their ratings of their trust in the speaker is based on a scale of 1 to 10, with 1 representing "Not at all" and 10 representing "Very much".

However, after conducting the pilot test, some issues were detected regarding the validity and

similarity of the voices as well as duration of the audio clips. Some independent raters reported that the audio clips all sound alike, even after the pitch and speaking rate have been digitally manipulated. This shows that the extent of manipulation was not significant enough for the listeners to detect any difference. However, if the pitch and speaking rate of the speakers in the audio clips were manipulated any more significantly, the voices would sound too robotic and unnatural. As this research is investigating the effects of vocal cues on trust at zero acquaintance, a realistic-sounding voice is vital to ensure that the listeners are giving accurate ratings as if the voices came from human sources. Therefore, some changes must be made with regards to the methodology. These changes are shown in Fig. 2.

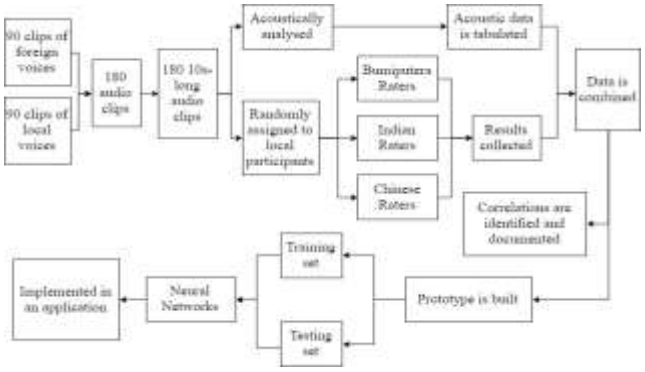

Fig. 2 A general overview of the investigation procedure

Instead of digitally manipulating the audio clips, the author has made the decision to collect a minimum of 180 audio clips. This number is selected because it is realistic and achievable yet hopefully enough to train and test the neural network which will be used in the final prototype. Half of these audio clips will be taken from the OMG-Emotion behavior dataset [46], and the other half will be sourced from YouTube. Since this area of research is new, there has yet to be any control group of "trustworthy voices" that exist. It can be said that this research will be creating its own dataset of "trustworthy voices" which may be released in the future. This dataset will potentially serve to support research purposes in the field of affective computing in the future.

The OMG-emotion dataset consists of a list of YouTube videos with several audio clips per video, each labelled with their own features. Some examples of features include start time, end time, emotion tags as well as several other acoustic characteristics like arousal and valence. However, in order to equally investigate the influence of accents (foreign vs local) on perceived trustworthiness, the dataset must contain audio clips from Malaysian speakers as well. As the OMG-Emotion behavior dataset does not consist of any Malaysian speakers, the audio clips of Malaysian speakers had to be sourced from YouTube. The dataset will have an equal distribution of local and foreign speakers, namely 90 from local speakers and 90 from foreign speakers. The speakers in each audio clip clearly do not know at that time of speaking that their voices would be used to measure trustworthiness. As such, there is no possibility of the voices being consciously manipulated at the point of speaking to form a "trustworthy voice".

These raw audio clips will be cropped to form 10-second-long audio clips. This is in accordance with the findings of [34] which indicates that a 10-second-long audio clip could be long enough to capture personality impressions. The 10-second-long audio clips will contain as little information about the context of the speech as possible and potentially consist of a few filler words such as "ummm" and "ahhh". Although the impact of lexical and semantic information of the voice cannot be eliminated [47], this approach will significantly reduce the impact of semantic information on the listener's ratings by only choosing audio clips that consists of words that do not have a context. These 10-second-long audio clips will be grouped into sets of 6. Each set of 6 will consist of a mixture of foreign and local accents as well as a variety of emotions.

These sets will be randomly assigned to local participants. An estimated amount of 60-80 Malaysian participants of various age groups are required to participate in this survey. This number is selected because it is realistic and achievable. The participants must contain an equal number of representatives from each ethnic group. These main ethnic groups consist of Bumiputra group as well as Chinese and Indian ethnic groups. As this investigation is targeted towards Malaysians, the participants must be Malaysian. To obtain as many participants as possible, the age range of the participants are not limited. A large number of people are encouraged to participate in order to increase the validity of this experiment and to reduce the impact of straight-liner judgements. The participants will have to state their age group as well as ethnic group after completing the survey.

The participants will be asked a series of questions regarding those audio clips through an online survey. The participants will be asked to rate their degree of trust in the speaker of each audio clip. In addition to that, the participants will also be questioned regarding the intelligibility of the words spoken, the emotions portrayed by the speaker, and the accent of the speaker. After rating one set, the same participant will also be able to rate other sets if he/she still has the capacity to do so. The questions asked in each set are similar, but the audio clips attached to each set are different. The audio clips can be replayed as many times as necessary. Due to the nature of this investigation, the volume and the device used will vary according to the individual. However, such a situation is realistic in the sense that

it mimics the natural way humans hear the voices of other individuals. Under natural circumstances, the voice of the speaker is varying in volume and clarity, depending on the environment and distance between the listener and the speaker. After the participants have submitted their responses for their respective sets, the responses will be analyzed.

Meanwhile, the series of 10-second-long audio clips will be subjected to acoustic analysis by using SoundGen and the acoustic characteristics of the audio clips will be recorded as well. Subsequently, the acoustic characteristics and the responses of the participants in the survey will be collected and tabulated alongside each other. After that, the correlation between the features of the audio clips and the ratings given by the listeners will be analyzed by using a correlation matrix. At the very least, a relationship is hoped to be identified between voice characteristics and perceived trustworthiness.

As a demonstration of this research project, a model will be produced. Due to the limited and relatively small sample size, a neural network model will be used to predict the degree of trustworthiness of a person's voice. The neural network model will first be trained with half of the labelled dataset, and the other half of the dataset will be used to test the neural network model.

IV. EXPECTED DELIVERABLES

This research aims to produce two deliverables. Firstly, this research aims to produce a detailed description of the voice characteristics that influence trustworthiness. Secondly, this research aims to produce a model that will be able to predict the degree of trustworthiness of an individual based solely on the voice. In order to achieve the objective, a neural network model is suggested to be used. This because neural network models can work significantly well even with a small sample size, as compared with other machine learning models such as random forest. This model is planned to utilize the TensorFlow platform to build the Artificial Neural Network model. The model will be trained with half of the labelled dataset, and the other half of the dataset will be used to test the model.

V. CONCLUSION

It is important to note that this research is still at the proposal stage. This research has yet to be carried out fully, and therefore is unable to deliver any conclusive results yet. However, this paper suggests a method towards conducting the research. The ultimate aim of the research is to investigate the association between vocal features and trustworthiness at zero acquaintance. In later stages of this research, this approach will be pilot tested and if the approach succeeds in producing a correlation, this approach will be used and run on a full scale. Otherwise, further research will have to be conducted to determine a suitable approach in order to conduct this investigation. Once a successful approach has been determined, the correlation between vocal features and trustworthiness at zero acquaintance will be documented. The data collected will then be used to build a neural network model which will be able to predict the degree of trustworthiness in a human's voice. This neural network model will be useful when validating the trustworthiness of the robotic voice when it is developed in the future. Ultimately, this research aims to contribute to the field of online tutoring and in particular, to aid in the creation of a more trustworthy robot tutor voice.

ACKNOWLEDGMENT

The work in this research are funded by grant number 304/PKOMP/6315137 and grant number 1001/PKOMP/8014001.

REFERENCES

[1] Trust, *Cambridge Advanced Learner's Dictionary and Thesaurus*. Cambridge University Press, 1999.

[2] A. G. Wooten, and J. C. McCroskey, "Student trust of teacher as a function of socio communicative style of teacher and socio-communicative orientation of student," *Communication Research Reports*, vol. 13, no. 1, pp. 94-100, 1996. Available: https://doi.org/10.1080/08824099609362075 doi: 10.1080/08824099609362075

[3] D. V. Maele, and M. V. Houtte, "The quality of school life: Teacher-student trust relationships and the organizational school context," *Social Indicators Research*, vol. 100, no. 1, pp. 85–100, Jan, 2010. Available: doi: 10.1007/s11205-010-9605-8

[4] A. Cook-Sather, "Authorizing students' perspectives: Toward trust, dialogue, and change in education," *Educational Researcher*, vol. 31, no. 4, pp. 3-14, 2002. Available: https://doi.org/10.3102/0013189X031004003 doi:10.3102/0013189X031004003

[5] A. Gregory, and M.B. Ripski, "Adolescent trust in teachers: Implications for behavior in the high school classroom," *School Psychology Review*, *37*(3), p.337, 2008.

[6] S. Lee, "The relations between the student–teacher trust relationship and school success in the case of Korean middle schools," *Educational Studies*, vol. 33, no. 2, pp. 209-216, 2007. Available: https://doi.org/10.1080/03055690601068477 doi: 10.1080/03055690601068477

[7] R. N. Stanulis, and D. Russell, "Jumping in trust and communication in mentoring student teachers," *Teaching and Teacher Education*, vo. 16, no. 1, pp. 65 – 80, 2000. Available: http://www.sciencedirect.com/science/article/pii/S0742051X99000414 doi:https://doi.org/10.1016/S0742-

051X(99)00041-4

[8] N. Ambady, M. A. Krabbenhoft, and D. Hogan, "The 30-sec sale: Using thin-slice judgments to evaluate sales effectiveness," *Journal of Consumer Psychology*, vol. 16 no., 1, pp. 4–13, 2006.

[9] M. Stirrat, and D. Perrett, "Valid facial cues to cooperation and trust: Male facial width and trustworthiness," *Psychological Science*, vol. 2, no. 3, pp. 349-354, 2010. Available: https://doi.org/10.1177/0956797610362647 (PMID:20424067) doi: 10.1177/0956797610362647

[10] D. A. Kenny, "Zero Acquaintance," http://davidakenny.net/ip/zeroa.htm Mediation, 2002. (David A. Kenny). Available: D. A. Kenny and L. Albright, "Accuracy in interpersonal perception: a social relations analysis," *Psychological Bulletin*, vol. 102, no. 3, p. 390, 1987.

[11] N. Eisenkraft, "Accurate by way of aggregation: Should you trust your intuition-based first impressions?" *Journal of Experimental Social Psychology*, vol. 49 no. 2, pp. 277-279, 2013. Available: http://www.sciencedirect.com/science/article/pii/S0022103112002235 doi:https://doi.org/10.1016/j.jesp.2012.11.005

[12] N. Ambady, and R. Rosenthal, "Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness" *Journal of personality and social psychology*, vol. 64, no. 3, p.431, 1993.

[13] M. A. Hecht, and M. LaFrance, "How (fast) can I help you? tone of voice and telephone operator efficiency in interactions 1," *Journal of Applied Social Psychology*, vol. 25, no. 23, pp. 2086–2098, 1995.

[14] N. Ambady, and H. M. Gray, "On being sad and mistaken: Mood effects on the accuracy of thin-slice judgments," *Journal of Personality And Social Psychology*, vol. 83, no. 4, p. 947, 2002.

[15] J. Fenwick, L. Barclay, and V. Schmied, "Chatting: An important clinical tool in facilitating mothering in neonatal nurseries," *Journal of Advanced Nursing*, vol. 33, no. 5, pp. 583–593, 2001. Available: doi: 10.1046/j.1365-2648.2001.01694.x

[16] N. Bos, J. Olson, D. Gergle, G. Olson, and Z. Wright, "Effects of four computer-mediated communications channels on trust development," *In Proceedings of The Sigchi Conference on Human Factors in Computing Systems*, pp. 135–140, 2002. New York, NY, USA: ACM. Available: http://doi.acm.org/10.1145/503376.503401 doi:10.1145/503376.503401

[17] L. Anolli, and R. Ciceri, "The voice of deception: Vocal strategies of naive and able liars," *Journal of Nonverbal Behavior*, vol. 21, no. 4, pp. 259–284, 1997.

[18] M. Zuckerman, R. S. DeFrank, J. A., Hall, D. T. Larrance and R. Rosenthal, "Facial and vocal cues of deception and honesty," J*ournal of Experimental Social Psychology*, vol. 15, no. 4, pp. 378-396, 1979. Available: http://www.sciencedirect.com/science/article/pii/0022103179900453 doi:https://doi.org/10.1016/0022-1031(79)90045-3

[19] C. Kirchhübel, and D. M. Howard, "Detecting suspicious behaviour using speech: Acoustic correlates of deceptive speech–an exploratory investigation," *Applied Ergonomics*, vol. 44, no. 5, pp. 694–702, 2013.

[20] T. Hatcher, "Research integrity: Ensuring trust in the academy," *Human Resource Development Quarterly*, vol. 16, no. 1, pp. 1–6, Mar, 2005. Available: doi: 10.1002/hrdq.1120

[21] J. Publishers, J. Perry and J. Kouzes, The J*ossey-Bass Reader on Nonprofit and Public Leadership*. John Wiley & Sons, 2009. Available: https://books.google.com.my/books?id=NAD8IXtTWrgC

[22] L. Qiu, and I. Benbasat, "Online consumer trust and live help interfaces: The effects of text-to-speech voice and three-dimensional avatars," *International Journal of Human–Computer Interaction*, vol. 19, no. 1, pp. 75-94, 2005. Available: https://doi.org/10.1207/s15327590ijhc1901_6 doi: 10.1207/s15327590ijhc1901_6

[23] S. Greenspan, D. Goldberg, D. Weimer, and A. Basso, "Interpersonal trust and common ground in electronically mediated communication," in *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, pp. 251–260, 2000. New York, NY, USA: ACM. Available: http://doi.acm.org/10.1145/358916.358996 doi: 10.1145/358916.358996

[24] M. D. Spector, and G. E. Jones, "Trust in the workplace: Factors affecting trust formation between team members," *The Journal of Social Psychology*, vol. 144, no. 3, pp. 311-321, 2004 Available: https://doi.org/10.3200/SOCP.144.3.311-321 (PMID: 15168431) doi:10.3200/SOCP.144.3.311-321

[25] L. Gilson, N. Palmer, and H. Schneider, "Trust and health worker performance: Exploring a conceptual framework using South African evidence," *Social Science & Medicine*, vol. 61, no. 7, pp. 1418-1429, 2005. Available: http://www.sciencedirect.com/science/article/pii/S027795360400646X (Building Trust and Value in Health Systems in Low- and Middle- Income Countries) doi: https://doi.org/10.1016/j.socscimed.2004.11.062

[26] C. Moorman, R. Deshpandé, and G. Zaltman, "Factors affecting trust in market research relationships," *Journal of Marketing*, vol. 57, no. 1, pp. 81–101, 1993. Available: http://www.jstor.org/stable/1252059

[27] A. Abdul-Rahman and S. Hailes, "Supporting trust in virtual communities," *In System sciences, 2000. Proceedings of the 33rd annual Hawaii international conference on* p. 9, 2000.

[28] C. M. Ridings, D. Gefen, and B. Arinze, "Some antecedents and effects of trust in virtual communities," *The Journal of Strategic Information Systems*, vol. 11, no. 3, pp. 271-295, 2002. Available: http://www.sciencedirect.com/science/article/pii/S0963868702000215 doi:https://doi.org/10.1016/S0963-8687(02)00021-5

[29] D. J. Kim, Y. I. Song, S. Braynov, and H. Rao, "A

multidimensional trust formation model in b-to-c e-commerce:a conceptual framework and content analyses of academia/practitioner perspectives," *Decision Support Systems*, vol. 40, no. 2, pp. 143-165, 2005. Available: http://www.sciencedirect.com/science/article/pii/S0167923604000107 doi: https://doi.org/10.1016/j.dss.2004.01.006

[30] M. K. O. Lee and E. Turban, "A trust model for consumer internet shopping," *International Journal of Electronic Commerce*, vol. 6, no. 1, pp. 75-91, 2001. Available: https://doi.org/10.1080/10864415.2001.11044227 doi: 10.1080/10864415.2001.11044227

[31] L. Albright, D. A. Kenny, and T. E. Malloy, "Consensus in personality judgments at zero acquaintance," *Journal of Personality and Social Psychology*, vol. 55, no. 3, p. 387, 1988).

[32] D. C. Funder and K. M. Dobroth, "Differences between traits: Properties associated with interjudge agreement," *Journal of Personality and Social Psychology*, vol. 52, no. 2, p. 409, 1987.

[33] D. A. Kenny, L. Albright, T. E. Malloy, and D. A. Kashy, "Consensus in interpersonal perception: Acquaintance and the big five," *Psychological Bulletin*, vol. 116, no. 2, p. 245, 1994.

[34] S. M. Mohammad and S. Kiritchenko, "Using nuances of emotion to identify personality" *Proceedings of ICWSM*, 2013.

[35] S. L. Lutfi, F. FernáNdez-MartíNez, J. M. Lucas-Cuesta, L. LóPez-LebóN, and J. M. Montero, "A satisfaction-based model for affect recognition from conversational features in spoken dialog systems," *Speech Communication*, vol. 55 no. 7-8, pp. 825–840, 2013.

[36] S. M. Hughes and M. A. Harrison, "Your cheatin' voice will tell on you: Detection of past infidelity from voice" *Evolutionary Psychology*, vol. 15, no. 2, 2017. .Available: https://doi.org/10.1177/1474704917711513 (PMID: 28580806) doi: 10.1177/1474704917711513

[37] P. Belin, B. Boehme, and P. McAleer, "The sound of trustworthiness: Acoustic-based modulation of perceived voice personality," *PLOS ONE*, vol. 12, no. 10, pp. 1-9, 2017. Available: https://doi.org/10.1371/journal.pone.0185651 doi: 10.1371/journal.pone.0185651

[38] I. Torre, L. White, and J. Goslin, "Behavioural mediation of prosodic cues to implicit judgements of trustworthiness," *In Proceedings of Speech Prosody*, pp. 816-820, 2016.

[39] H. S. Cheang and M. D. Pell, "The sound of sarcasm," *Speech Communication*, vol. 50, no. 5, pp. 366-381, 2008. Available: http://www.sciencedirect.com/science/article/pii/S0167639307001884 doi:https://doi.org/10.1016/j.specom.2007.11.003

[40] B. L. Smith, B. L. Brown, W. J. Strong, and A. C. Rencher, "Effects of speech rate on personality perception," *Language and Speech*, vol. 18, no. 2, pp. 145-152, 1975. Available: https://doi.org/10.1177/002383097501800203 (PMID:1195957) doi: 10.1177/002383097501800203

[41] M. Imhof, "Listening to voices and judging people" *International Journal of Listening*, vol. 24, no. 1, pp. 19-33, 2010. Available: https://doi.org/10.1080/10904010903466295 doi:10.1080/10904010903466295

[42] J. Trouvain, S. Schmidt, M. Schröder, M. Schmitz, and W. J. Barry, "Modelling personality features by changing prosody in synthetic speech," 2006.

[43] D. Birch and J. McPhail, "Does accent matter in international television advertisement?," *International Journal of Advertising*, vol. 18, no. 2, pp. 251–268, 1999.

[44] P. H. Andersen, and R. Kumar, "Emotions, trust and relationship development in business relationships: A conceptual model for buyer–seller dyads," *Industrial Marketing Management*, vol. 35, no. 4, pp. 522–535, 2006.

[45] E. Tsankova, G. Möllering, A. Kappas, E., Krumhuber, A. Aubrey, A. Manstead, and P. Rosin, "Visual and paralinguistic cues, confidence, and perceived trustworthiness," *In Proceedings of the 6th workshop on Trust within and between Organizations*, pp. 1–15, 2012.

[46] P. Barros, N. Churamani, E. Lakomkin, H. Siqueira, A. Sutherland, and S. Wermter, "The omg-emotion behavior dataset," 2018. Available: doi: arXiv preprint arXiv:1803.05434

[47] D. J. Hirst, "Form and function in the representation of speech prosody," *Speech Communication*, vol. 46, no. 3-4, pp. 334–347, 2005.