Voicing Trust: An Acoustic Analysis of Trustworthiness in Automated Customer Service Interfaces

SOBAN ASIF, University of Twente, The Netherlands

The advent of synthetic voices in automated customer service systems has revolutionized the way businesses interact with their clientele. However, the success and effectiveness of these interactions largely hinges on the trustworthiness of the synthetic voice. This research aims to dissect the combination of acoustic parameters that contribute to a voice's trustworthiness, focusing on gender, pitch, speaking rate, and the duration of pauses between punctuation marks. Through the development of a graphical user interface that allows users real-time manipulation of these acoustic features, they attempt to design a trustworthy synthetic voice. The study will investigate how these specific acoustic qualities are selected to enhance the credibility of the voice. The research employs a mixed-methods approach, combining quantitative analysis of the selected acoustic features with qualitative feedback to understand the nuances of these settings, the most valuable features contributing to trustworthiness and understanding how the co-design process could be improved. Findings indicate a preference for low pitch values and moderate speaking rates, with strategic use of pauses enhancing communication effectiveness. The results provide valuable guidelines for designing more trustworthy and engaging synthetic voices, improving user satisfaction and the effectiveness of automated customer service systems.

Additional Key Words and Phrases: Synthetic voices, Trustworthiness, Automated customer service, Voice Design

1 INTRODUCTION

In the evolving landscape of customer service, the integration of synthetic voices within automated systems has become increasingly prevalent. An example of this trend is the AiCall startup at the University of Twente [11], which exemplifies the growing use of synthetic voices in the customer service sector. The auditory representation of these systems plays a critical role in shaping user trust and overall experience. The real-time nature of chat services has transformed customer service into a two-way communication with significant effects on trust and satisfaction [9], This project aims to develop a user interface that allows for the manipulation of basic acoustic features—such as gender, pitch, speaking rate, and the duration of pauses between punctuation marks. This interface will enable users to design a synthetic voice that they deem trustworthy and allow them to flesh out why they designed it that way. The use of the interface would also allow different customer service companies to cater to their specific customers.

While these features have been individually explored in prior studies [25] with respect to trust, their combination in the context of customer service has not been thoroughly investigated. Exploring the interplay of these acoustic features is essential, as individual features often do not yield linear results, and research findings can be inconclusive or even contradictory [25]. The interface will serve as a research platform to investigate how these acoustic parameters collectively influence the trustworthiness of synthetic voices, enabling a co-design process where users actively participate in shaping and refining the acoustic qualities to enhance trustworthiness.

The gender of the synthetic voice is a critical factor in this context. Research indicates that gendered voices can influence user preferences and perceptions of competence and warmth which in turn influence the trustworthiness of the synthetic voice [16]. Pitch and speaking rate are also significant, as they can affect assessments of authority and agreeableness. For instance, a study by Belin et al. (2017) demonstrated that certain acoustic modulations could significantly alter the trustworthiness of a voice [3]. Similarly, pauses within speech play a crucial role in communication. Strategic use of pauses can enhance speech intelligibility and convey thoughtfulness, potentially increasing the trustworthiness of the system [20].

However, the lack of comprehensive studies on the combined effects of these acoustic features presents a significant challenge for designers of synthetic voices. Without clear guidelines, businesses risk deploying synthetic voices that may inadvertently diminish user trust, thereby undermining the effectiveness of automated customer service systems. This problem is compounded by the absence of user-friendly tools that allow for real-time manipulation and testing of these acoustic features to identify the most trustworthy configurations.

To address this gap, our research is guided by the following primary research question (RQ):

RQ: How would a trustworthy voice in an automated customer service context sound based on the manipulation of its acoustic features through a basic user interface?

To explore the primary research question in depth, the following sub-research questions (SRQ) were derived on a feature level and formulated:

- **SRQ1:** How does the selected **gender** of a synthetic voice contribute to its trustworthiness in customer service interactions?
- **SRQ2:** What role does the manipulation of **pitch** play in building a trustworthy synthetic voice, and what pitch ranges are associated with higher levels of trust?
- **SRQ3:** How does the manipulation of **speaking rate** influence the trustworthiness of a synthetic voice, and what rates are preferred by users?
- **SRQ4:** What is the effect of manipulating the **duration of pauses** between punctuation marks of sentences on the trustworthiness of synthetic voices?
- **SRQ5:** What features are deemed the most important when constructing a trustworthy synthetic voice?

TScIT 41, July 5, 2024, Enschede, The Netherlands

- **SRQ6:** What features are deemed the least important when constructing a trustworthy synthetic voice?

By addressing these sub-questions in combination with the other attributes, the research aims to construct a detailed profile of a trustworthy synthetic voice, providing actionable insights for the design of automated customer service systems.

2 RELATED WORKS

The study of synthetic voices and their trustworthiness in customer service applications has been a growing area of interest in recent decades. Previous research has focused on various acoustic features individually, providing a foundation for understanding how these features influence perceptions of trust and credibility.

2.1 Defining Trustworthiness

Defining trustworthiness in the context of synthetic voices is crucial for understanding how various acoustic features influence users' design process of their trustworthy synthetic voice.

To establish a clear operational definition of trustworthiness for this study, a thorough literature review was conducted. Trustworthiness in human-computer interactions often encompasses the following attributes as defined by Mayer, Davis, and Schoorman's (1995) integrative model of organizational trust [15]:

- **Ability:** The perceived competence and skills of the synthetic voice to understand and respond appropriately to user queries. This includes aspects such as the intelligence and effectiveness of the voice in providing accurate and useful information [5, 15].
- **Integrity:** The perceived honesty and adherence to principles of fairness and truthfulness by the synthetic voice. This involves the avoidance of misleading or deceptive information and ensuring that the voice is sincere and truthful [13, 15].
- **Benevolence:** The perception that the synthetic voice has the user's best interests at heart and is supportive and empathetic. This includes the synthetic voice showing concern for the user's welfare and acting in a way that benefits the user [15].

By integrating insights from existing research, the following working definition of trustworthiness was established for this study:

A trustworthy synthetic voice is one that a user deems as able, having integrity, and benevolent, thereby fostering confidence and comfort in interactions with automated customer service systems.

This definition guided the design and evaluation of the user interface and the manipulation of acoustic features. It also provided a framework for interpreting the results of the study, ensuring that the identified optimal combinations of acoustic features are consistent with a user's trust in the synthetic voice.

2.2 Gender and Trustworthiness

Gendered synthetic voices have a significant impact on user perceptions in customer service interactions. For trust interactions, research did not find conclusive consistent differences in trust judgement towards men and women [25] [19] [22].

Lopatovska et al. (2021) investigated the effects of gendered voices on personality perceptions of conversational user interfaces, revealing that users generally preferred female voices for casual interactions, while male voices were favored in professional and stressful contexts involving trust [14]. Further research by Lee et al. (2000) found that synthetic male voices had a greater impact on users' decisions and were perceived as more socially attractive and trustworthy than synthetic female voices [12]. Mullennix et al. (2003) also found that messages delivered by male synthetic voices were rated as more favorable, persuasive, and trustworthy in terms of persuasive appeal compared to those delivered by female synthetic voices [18]. By contrast, Todorov et al. (2014) determined that the first impression of vocal attractiveness in male voices relates to perceived strength and authoritativeness, whilst in females, vocal attractiveness relates to perceived warmth and trustworthiness [16]. Additionally, gender stereotypes can influence perceived trustworthiness, where users might trust voices that align with expected gender roles in specific contexts [17].

The integration of these findings into the co-design process emphasizes the complex interplay between gender, context, and user interaction. Designing synthetic voices with these considerations in mind can optimize trustworthiness and user satisfaction, making the co-design approach vital for effective automated customer service systems.

2.3 Pitch, Speaking Rate, Pauses, and Trustworthiness

Pitch is a critical factor in voice trustworthiness. Studies have shown that lower-pitched voices are generally perceived as more trustworthy and competent [10, 16]. For instance, Shiramizu et al. (2022) demonstrated that lower-pitched voices were perceived as more dominant, affecting the trustworthiness of AI conversational agents [23]. Similarly, the work of Belin et al. (2017) highlighted that pitch modulation can significantly alter a listener's perception of trustworthiness [3]. Apple et al. (1979) discovered that higher pitched voices were judged as less truthful and less emphatic affecting the trustworthiness negatively [2]. A study by Oleszkiewicz et al. (2017) was conducted with blind and sighted participants. Researchers had participants rate voices for trustworthiness, competence, and warmth and found that both sets of participants rated lower-pitched men's voices as more competent and trustworthy while higher-pitched voices were rated as warmer [21]. Similarly, Elkins and Derrick (2013) found that higher pitched voices lowered perceptions of trust in the agent, but only at the beginning; the effect flattened out over time [4]. These findings suggest that careful modulation of pitch is essential in the design of synthetic voices to enhance user trust.

The speaking rate of a synthetic voice also influences its perceived trustworthiness. The relationship between empathy and trustworthiness lies in the user's perception that the voice understands and responds appropriately to their needs, thus fostering trust. Faster speaking rates are often associated with greater competence but can reduce perceived empathy, which in turn may affect trustworthiness. Conversely, slower rates may enhance perceptions of warmth and thoughtfulness which can enhance trust, all the while they may also be perceived to be less competent and therefore untrustworthy. Research by Apple et al. (1979) indicated that an optimal speaking

rate is crucial for balancing competence and empathy, which are both components of trustworthiness [2]. A fast speaking rate was found to be a feature of charismatic and trustworthy persuasive speakers [7, 8] meanwhile slower speech rates predicted greater trusting behaviour in other studies [26].

The strategic use of pauses in speech plays a crucial role in communication effectiveness and trustworthiness. Pauses can enhance speech intelligibility and provide listeners with time to process information, thereby increasing the perceived thoughtfulness and reliability of the speaker. Ohta et al. (2014) found that pauses and silences added at natural breaks within sentences improved comprehension of the information presented by an agent as well as the naturalness and trustworthiness of the agent's synthetic voice [20]. In contrast, Jeong et al. (2019) found that likeability when vocal fillers were used depended on the context, specifically better for social situations rather than in a service context [6]. Research by Skantze et al. (2013) indicated that users utilize the robot's gaze and pauses to manage the flow of interaction and that these cues affect user behavior, which is essential for building trust in human-robot interaction [24]. These findings highlight the importance of the strategic use of pauses in designing trustworthy synthetic voices.

2.4 Combining Acoustic Features for Trustworthiness

While individual acoustic features have been extensively studied, their combined effect on trustworthiness in synthetic voices remains under-explored. Torre et al. (2021) suggested that it is unlikely that the relationships between trust attributions and, for example, speech rate or pitch range are strictly linear. Additionally, rather than individual vocal characteristics, it is more likely to be a combination of these acoustic features that determine the participant's assessment of trustworthiness [25]. The acoustic features are also deeply correlated with each other. Devers et al. (2024) indicate that people associate the synthetic voice with gender atypical characteristics such as pitch [17]. Mullennix et al. (2003) also found that synthetic male voices were associated with speaking slowly [18]. This highlights the need for comprehensive studies that investigate how combinations of gender, pitch, speaking rate, and pauses can be manipulated to optimize trustworthiness in synthetic voices. Understanding these nuances is critical for designing effective automated customer service systems that users can trust.

This research aims to fill this gap by providing a detailed analysis of how these acoustic features can be combined and manipulated to build a trustworthy synthetic voice. The development of a user interface for real-time manipulation of these features will facilitate a deeper understanding of their interactive effects, enabling a co-design process where users actively participate in shaping and refining the acoustic qualities, and contribute to the design of more effective automated customer service systems.

3 METHODOLOGY

To investigate the impact of acoustic features on the trustworthiness of synthetic voices in automated customer service systems, an online platform was created in order to facilitate participants with designing their trustworthy voice and in the process providing both qualitative and quantitative data to comprehensively understand user preferences and design parameters. The online platform was pilot tested by two users before being released for participants.

3.1 Participants

Participants were recruited through online advertisements and email invitations. The criteria for participation included being fluent in English. A total of 20 participants were recruited, consisting of a diverse demographic in terms of age, gender, and background.

Key ethical considerations included:

- **Informed Consent:** Participants provided informed consent before participating in the study.
- **Anonymity and Confidentiality:** Participant data were anonymized, and confidentiality was maintained throughout the research process.
- **Voluntary Participation:** Participation was voluntary, and participants could withdraw from the study at any time without any consequences.

3.2 Measurements

3.2.1 Voice Manipulation Interface (Quantitative Data): Participants were allowed to create their trustworthy voice through the interface settings section (shown in Figure 1) which included the following adjustable parameters:

- **Gender:** Options included male and female.
- **Speaking Rate:** Ranges from Very Low to Very High, with fixed corresponding values: Very Low (0.67), Low (0.8), Medium (1.0), High (1.25), and Very High (1.5). A speaking rate of 1.0 represents the normal, native speed of the specific voice and the corresponding values are just multiplications of the normal native speeds. This range ensures clarity and comprehensibility of the synthesized speech, allowing users to find a comfortable and understandable speed to create a trustworthy voice. A slider is used to adjust the speaking rate.
- **Pitch:** Adjustments range from Very Low to Very High, with values from -10.0 to 10.0. A pitch of 0.0 is the original pitch of the voice. A value of 10.0 increases the pitch by 10 semitones (approximately one octave), making the voice significantly higher, while -10.0 decreases the pitch by 10 semitones, making it significantly lower. This range provides flexibility in modifying the voice to suit different trust requirements while maintaining the natural quality of the voice. A slider is used to adjust the pitch.
- **Pause Duration:** Ranges from Low to High with values: Low (150ms), Medium (500ms), and High (800ms). The pauses are inserted at punctuation marks(full-stop, comma, exclamation point). This range helped simulate natural speech patterns.

These voice settings were encoded as integers and saved for subsequent quantitative analysis.

3.2.2 Questionnaire (Qualitative Data): The questionnaire consisted of the following open-ended questions designed to understand participant views and rationales regarding the trustworthiness of the synthetic voice they created, as well as suggestions for improving the co-design process:

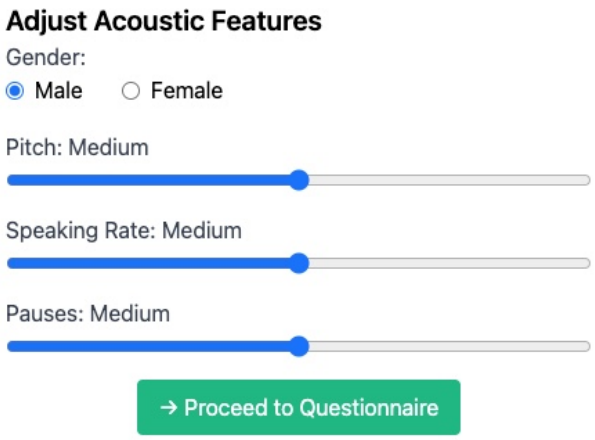

Fig. 1. Voice Settings

(1) Which acoustic feature(s) (speaking rate, pitch, gender, pauses) do you think contributed most to a trustworthy voice?
(2) Which acoustic feature(s) (speaking rate, pitch, gender, pauses) do you think contributed least to a trustworthy voice?
(3) Why do you think the specific gender setting you selected contributed to a trustworthy voice?
(4) Why do you think the specific pitch setting you selected contributed to a trustworthy voice?
(5) Why do you think the specific speaking rate setting you selected contributed to a trustworthy voice?
(6) Why do you think the specific pauses setting you selected contributed to a trustworthy voice?
(7) Overall, how satisfied are you with the synthetic voice you created?
(8) How could the co-design process of the synthetic voice be improved?
(9) Do you have any additional comments or suggestions?

These questions were crafted to elicit detailed responses about the participants' thought processes and the effectiveness of each acoustic feature in contributing to trustworthiness.

3.3 Task and Procedure

To facilitate the research, a web-based user interface was developed, enabling the entire study to be conducted online. This approach provided the convenience of automatic data collection, allowed participants to actively design their trustworthy synthetic voice, and simplified the process of recruiting participants.

Participants began by reading and completing the informed consent form. Once consent was obtained, they were directed to the voice manipulation interface (see Appendix B.1), where they received instructions on the task. The main task involved creating a trustworthy voice by manipulating acoustic features which were gender, pitch, speaking rate, and pause duration within an example chat-bot conversation scenario.

Participants were presented with a simulated chat conversation designed to mimic interaction with a customer service bot. The chat tree was designed to feature a balanced conversation where trust as defined in related works was a key element of the conversation. The chat interface and voice parameter settings (shown in Figure 1) were displayed. Participants could choose between discrete binary options for their messages, and the bot would reply accordingly, simulating a conversation. The bot's responses were based on the example chat tree and the currently selected voice settings. The interface included additional features, such as the ability to reset the chat interaction and replay any bot message with different voice parameter settings.

After designing their trustworthy synthetic voice using the voice manipulation interface, participants proceeded to a questionnaire aimed at gathering qualitative data on their experiences and preferences. Upon completion of the questionnaire, participants were thanked and debriefed.

3.3.1 *Technologies Used.*

- **Website Framework:** The website was built with Flask, a Python web framework, for its rapid development capabilities within the limited timeframe.
- **Database:** PostgreSQL was employed to securely store participant data.
- **Speech Synthesis:** The Google Cloud Text-to-Speech API was integrated into the website. Among the voices offered by the API, the Journey US voice was selected for its suitability in creating engaging conversational agents, leveraging the latest advancements in AudioLM technology [1]. The API also facilitated adjustments to the speaking rate, pitch, and gender of the synthetic voice. Additionally, a custom function was implemented to modify the duration of pauses between punctuation marks in sentences.
- **Data Analysis:** For the quantitative data analysis, Python libraries such as Matplotlib, Pandas, and Seaborn were utilized. These tools enabled the automatic retrieval of data from the database, comprehensive analysis, and visualization of the results.

3.4 Analysis

3.4.1 Quantitative Analysis. To systematically analyze the impact of acoustic features on the trustworthiness of synthetic voices, several quantitative analysis techniques were employed. These techniques include descriptive statistics and correlation analysis, which together provide a comprehensive understanding of how different voice parameters influence user trust.

Descriptive statistics were used to summarize and describe the basic features of the data collected. Key metrics calculated include:

- **Mean and Mode:** The mean (average) and mode (most frequent value) were calculated for the voice parameters (pitch, speaking rate, and pause duration) and gender, respectively. The mean provides a central value for the acoustic features, offering insight into the typical setting used by participants. The mode shows the most commonly selected setting for the synthetic voice.

- **Standard Deviation and Variance:** These measures of dispersion assess the variability in the voice parameter settings. A high standard deviation indicates a wide range of values, suggesting diverse participant preferences for that parameter. Conversely, a low standard deviation indicates that most participants selected similar values, suggesting a consensus in preference.
- **Frequency Distributions:** Frequency distributions were created for categorical variables (such as gender) and for binned continuous variables (such as pitch and speaking rate). These distributions visualize how often each value or range of values occurred in the dataset, helping identify common patterns and outliers.

Correlation analysis was conducted to examine the relationships between different acoustic features:

- **Spearman Rank Correlation Coefficient:** Spearman correlation, suited for ordinal or non-normally distributed data, measures the strength and direction of monotonic relationships between acoustic features (gender, pitch, speaking rate, and pause duration). It assesses whether changes in one feature tend to correspond with changes in another feature, such as gender influencing pitch variations in the synthetic voice.

This correlation analysis helped identify which voice parameters tend to vary together, providing insights into how changes in one parameter might influence others. Understanding these relationships is crucial for optimizing the overall voice design, as it allows researchers to anticipate the effects of adjusting multiple parameters simultaneously.

3.4.2 Qualitative Analysis. The qualitative data obtained from the open-ended questionnaire responses were analyzed using inductive thematic analysis. Additionally, functions were written to identify the most valuable and least valuable features as mentioned by participants.

To determine the most and least valuable features, specific questions in the questionnaire asked participants to identify which features contributed the most and least to the trustworthiness of their voice. A Python function was written to normalize the responses and count the number of times each feature was mentioned in the responses.

Thematic analysis is a method for identifying, analyzing, and reporting patterns (themes) within data. The following steps were followed in the thematic analysis:

(1) **Generating Initial Codes:** Significant phrases or sentences in the responses were highlighted to generate initial codes. These codes represent meaningful segments of the data related to the research questions.
(2) **Searching for Themes:** The initial codes were examined to identify overarching themes.
(3) **Reviewing Themes:** The identified themes were reviewed and refined to ensure they accurately represent the data and are distinct from each other. Some themes were merged or split as needed.
(4) **Defining and Naming Themes:** Clear definitions and names were assigned to each theme to succinctly capture their essence and scope.

4 RESULTS

4.1 Quantitative Results of Synthetic Voice Settings

The synthetic voice settings were encoded as follows for analysis:

- Speaking rate and pitch were categorized as Very Low (0.0), Low (1.0), Medium (2.0), High (3.0), and Very High (4.0).
- The duration of pauses was encoded as Low (0.0), Medium (1.0) and High(2.0).

In Table 1, we can observe that the mean speaking rate is 1.95, indicating that the average speaking rate falls mostly at the Medium(2.0) settings with a very slight lean towards Low(1.0) rather than High(3.0). Pitch leans more towards a Low(1.0) value with a mean of 1.85. The standard deviation for pitch is 0.812728, which is higher than that of speaking rate(0.510418) and pause duration(0.7451), indicating greater variability in pitch settings among participants. This is supported by the distribution graphs shown in Figure 2. It can be seen that participants most commonly selected the Medium(2.0) settings for speaking rate while there are more spread out distributions for the pitch selection and duration of pauses selection. Table 2 and Figure 2 show the gender distribution of the synthetic voices. The gender distribution leans slightly towards the male choice, with 13 occurrences out of 20.

Figure 3 presents the Spearman Correlation Matrix, which illustrates the relationships between different acoustic features. Notably, pitch and gender show a high positive correlation with a value of 0.52 while pitch and speaking rate seem to have the lowest correlation with a value of 0.0048.

Descriptive Statistics	speaking_rate	pitch	pause
count	20	20	20
mode	Medium (2.0)	Low (1.0)	Medium (2.0)
frequency	15	8	9
mean	1.95000	1.85000	1.15000
standard deviation	0.510418	0.812728	0.74516
variance	0.260526	0.660526	0.555263
min	1.00	1.00	0.00
25%	2.00	1.00	1.00
50%	2.00	2.00	1.00
75%	2.00	3.00	2.00
max	3.00	3.00	2.00

Table 1. Descriptive Statistics of Acoustic Features

Descriptive Statistics	gender
count	20
mode	male
frequency	13

Table 2. Descriptive Statistics of Gender

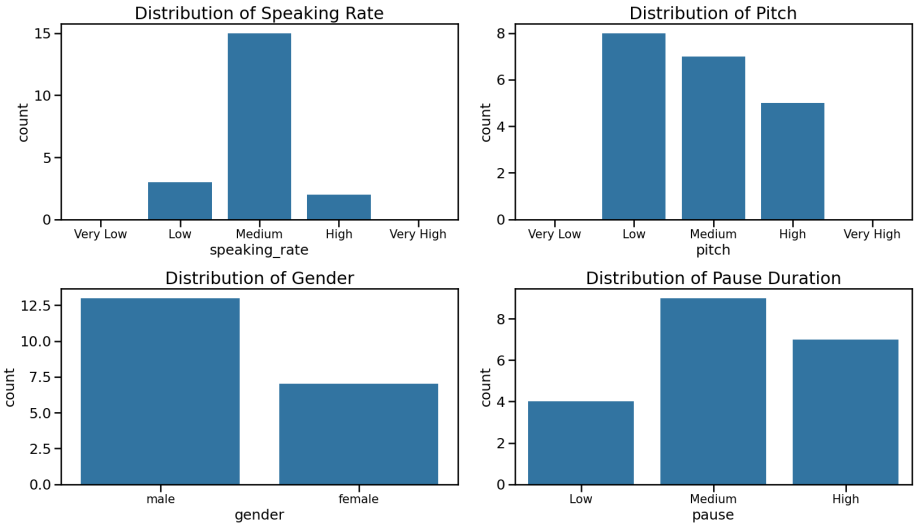

Fig. 2. Voice Settings Distribution

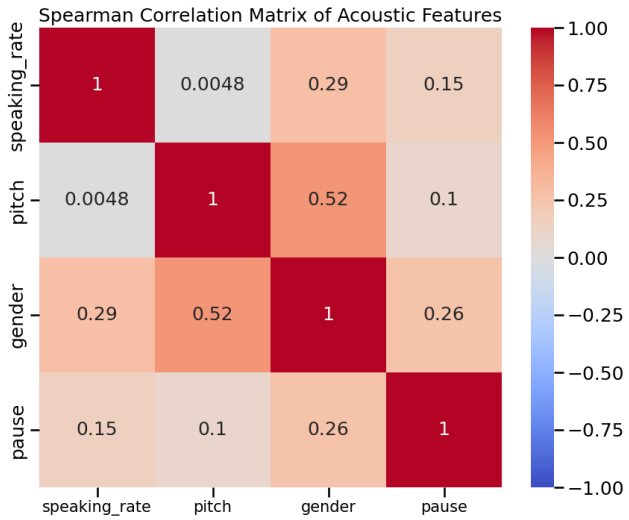

Fig. 3. Voice Settings Distribution

4.2 Qualitative Results of Questionnaire Responses

Using the count feature mentions function, it was found that the pitch of the synthetic voice was considered the most valuable feature, while the duration of pauses was deemed the least valuable.

Thematic analysis of the participant responses revealed several key themes, providing insights into why specific voice settings were chosen and how these settings contributed to the trustworthiness of the synthetic voice. The identified themes for each feature are as follows:

- **Impact of Gender:**
  - Participants selecting female voices often described them as a "gentler, nurturing synthetic voice."
  - Participants selecting male voices often described them as a "calming, capable synthetic voice."
- **Role of Pitch:**
  - Lower and medium pitch voices were associated with the theme "pleasant and mature synthetic voice."
  - Higher pitch voices were associated with the theme "friendlier, alluring voice."
- **Significance of Speaking Rate:**
  - Lower and medium speaking rates were associated with the theme "calm and collected synthetic voice."
  - Higher speaking rates were associated with the theme "fluid, competent voice."
- **Importance of Pauses:**
  - Shorter pauses were linked to the theme "help provided quickly in conversation."
  - Medium and longer pauses were linked to the theme "thoughtful, natural, honest pace of conversation."

5 DISCUSSION

The objective of this research was to determine how different acoustic features of synthetic voices could be designed to influence their trustworthiness in an automated customer service context. The results provide valuable insights into user preferences and the interplay of gender, pitch, speaking rate, and pauses in shaping trustworthy synthetic voices.

5.1 Answer to the Research Question (RQ)

RQ: *How would a trustworthy voice in an automated customer service context sound based on the manipulation of its acoustic features through a basic user interface?*

The findings suggest that a trustworthy synthetic voice is characterized by specific configurations of acoustic features that resonate with user inputs on reliability, benevolence and competence. These configurations were derived from both quantitative and qualitative data, offering a comprehensive understanding of user preferences. To comprehensively address this research question, we broke it down into several sub-research questions (SRQs) focusing on individual acoustic features. By analyzing the responses to these SRQs, we can build a detailed understanding of what makes a synthetic voice trustworthy in this context.

5.2 Analysis of Sub-Research Questions (SRQs)

SRQ1: *How does the selected gender of a synthetic voice contribute to its trustworthiness in customer service interactions?*

The gender distribution data, presented in Table 2, reveals a slight preference among participants for male synthetic voices. This preference aligns with previous studies indicating that male voices are often perceived as more persuasive and competent [12, 14, 18]. The thematic analysis reinforces this finding, with participants describing male voices as "calming" and "capable." Conversely, participants who preferred female synthetic voices characterized them as "gentler" and "nurturing", which aligns with research by Todorov et al. (2014) [16]. It is also essential to consider that gender stereotypes can influence perceived trustworthiness, with users potentially favoring voices that match expected gender roles in our customer service context [17]. This dichotomy underscores the complex role of gender in shaping trust, suggesting that different contexts may require different gendered voices to maximize trustworthiness.

SRQ2: *What role does the manipulation of pitch play in building a trustworthy synthetic voice, and what pitch ranges are associated with higher levels of trust?*

Quantitative results show that pitch settings had the highest variability (standard deviation of 0.812728), suggesting diverse user preferences. The qualitative data with thematic analysis revealed that lower and medium pitch voices were associated with the theme "pleasant and mature," indicating a preference for these ranges in fostering trust. High pitch voices were linked to being "friendlier" and "alluring," suggesting that while they may be approachable, they might lack the maturity associated with trustworthiness. These findings are consistent with previous studies that identified lower pitches as more authoritative and trustworthy [3, 10, 16, 23].

Pitch modulation significantly influences perceived trustworthiness in synthetic voices. The preference for lower and medium pitches suggests they convey stability and maturity, enhancing trust. This is particularly relevant in customer service interactions where users seek reliable and competent responses. However, higher pitches may be advantageous in contexts emphasizing approachability and warmth. Thus, the optimal pitch varies depending on specific desired interaction style and contexts.

SRQ3: *How does the manipulation of speaking rate influence the trustworthiness of a synthetic voice, and what rates are preferred by users?*

The speaking rate data indicates a strong preference for medium rates, with a mean value of 1.95 and very low variability (0.260526). The qualitative themes identified lower and medium speaking rates as "calm and collected," aligning with user contributions to the co-design process. In contrast, higher speaking rates were associated with being "fluid" and "competent," which may be advantageous in fast-paced contexts but could potentially compromise reliability. This supports the notion that moderate speaking rates strike a balance between efficiency and clarity, essential for trust [25]. The preference for medium speaking rates suggests that users favor a balanced approach. Medium rates seem to offer a compromise, providing enough speed to convey competence and efficiency without sacrificing the clarity and empathy necessary for trust. This balance is essential in customer service interactions, where both understanding and efficiency are critical.

SRQ4: *What is the effect of manipulating the duration of pauses between punctuation marks of sentences on the trustworthiness of synthetic voices?*

The duration of pauses showed a mean value of 1.15, indicating a preference for medium to slightly high pauses. Qualitative analysis revealed that medium and longer pauses were perceived as "thoughtful" and indicative of a "natural, honest pace," enhancing the trustworthiness of the synthetic voice. Shorter pauses, while associated with promptness, might not convey the same level of consideration and deliberation valued in customer service interactions. These findings are aligned with previous research emphasizing the role of pauses in enhancing speech intelligibility and perceived thoughtfulness [20]. The insights from the findings and related works suggest that strategic manipulation of pause duration is crucial in designing trustworthy synthetic voices in customer service interactions, where professionalism and clarity are paramount.

SRQ5: *What features are deemed the most important when constructing a trustworthy synthetic voice?*

The count feature mentions function identified pitch as the most valuable feature. This highlights the critical role of pitch in influencing user trust in synthetic voice trustworthiness. The variability in pitch preferences additionally underscores the need for customizable pitch settings in synthetic voice interfaces to cater to diverse user preferences.

The Spearmen Correlation Matrix shown in Figure 3 also indicated a high positive correlation between pitch and gender with a value of 0.52, suggesting that these two features are often interlinked when users are creating their trustworthy voice. This alligns with previous research that indicates that people associate the synthetic voice with gender atypical characteristics such as pitch and that these features are deeply tied together [17, 25]. Therefore, in the context of customer service interactions, it would be crucial to allign the pitch selection with the gender selection to enhance the trust of the created synthetic voice.

SRQ6: *What features are deemed the least important when constructing a trustworthy synthetic voice?* According to the count feature analysis, the duration of pauses was identified as the least valuable feature. However, this observation does not diminish the significance of pauses in synthetic voice design. Instead, it highlights that while pauses play a supporting role, features such as pitch and speaking rate have a more immediate and pronounced impact on perceived trustworthiness for automated customer service interactions.

Pauses contribute to the naturalness and cadence of speech, allowing listeners time to process information and enhancing overall comprehension. Despite being less emphasized in the analysis, the strategic use of pauses can still significantly influence how trustworthy the synthetic voice is.

5.3 Implications for Design

The insights gained from this research have practical implications for the design of synthetic voices in automated customer service systems. By enabling customization of pitch, speaking rate, and pause duration, designers can create more trustworthy and user-friendly synthetic voices. Moreover, understanding the nuanced preferences for gendered voices with its strong ties to pitch, a carefully selected moderate speaking rate and strategic use of pauses can help tailor synthetic voices to specific customer demographics and contexts, enhancing overall user experience and satisfaction.

5.4 Improvements and Recommendations

Based on participant feedback and suggestions, several areas for improvement and recommendations for enhancing the synthetic voice system have been identified:

Naturalness and Personalization: Participants expressed a desire for a more natural sounding voice, suggesting enhancements such as incorporating breaths between statements and reducing robotic intonation. Personalizing responses by addressing users by name was also suggested to improve engagement and user experience.

Language and Comfort Statements: Adding support for multiple languages was recommended to broaden accessibility. Incorporating comforting statements during interactions, such as acknowledging and apologizing for issues, was also seen as a way to enhance user satisfaction and trust in the synthetic voice.

Future Directions: Participants highlighted the potential for future research and development with more advanced AI voice assistants like ChatGPT, which could mitigate the uncanny valley effect and offer more human-like intonation and informal speech options.

These recommendations aim to make the synthetic voice system more natural, personalized, and versatile, catering to diverse user needs and preferences while enhancing overall usability and satisfaction in designing their trustworthy voice.

5.5 Limitations and Future Research

While the study provides valuable insights, it is limited by the sample size and the specific context of customer service. Future research could explore these acoustic features in different contexts and with larger, more diverse participant groups. Additionally, investigating the long-term impact of synthetic voice interactions on user trust could provide deeper insights into designing more effective automated systems. Moving forward, several avenues for further research and development could be explored.

5.5.1 Comparison of Trustworthy Voices. Building on the identification of an average trustworthy voice from the current research, future studies will involve a comparative analysis. Participants will interact with two distinct voices—one identified as the average trustworthy voice and another as untrustworthy—alongside their own created synthetic voice. This comparison will allow for comprehensive trustworthiness ratings across different voices. Conducting correlation and regression analyses on the trustworthiness ratings of the three voice types (average trustworthy, untrustworthy, and created trustworthy voice) will provide deeper insights. Relationships between voice parameters (such as pitch, speaking rate, and pauses) and perceived trustworthiness can be explored further. This analysis aims to identify specific voice characteristics that significantly influence trust perceptions, contributing to the refinement of synthetic voice design.

By executing these planned research directions, the synthetic voice system can evolve to better meet user expectations and enhance its effectiveness across various applications.

6 CONCLUSIONS

Synthetic voices are pivotal in modern automated customer service, where trustworthiness plays a critical role in user satisfaction and effectiveness of the technology. This study explored how gender, pitch, speaking rate, and pauses can be manipulated to design trustworthy synthetic voices. Through quantitative analysis and qualitative feedback, it was found that low pitch values and moderate speaking rates are generally preferred, with strategic use of pauses enhancing communication effectiveness.

Participants attributed characteristics of empathy to female voices and authority to male voices, highlighting the role of gender in synthetic voice design. Pitch was identified as a key factor influencing perceptions of professionalism and approachability. The study underscores the importance of balancing naturalness and clarity in voice design to optimize user trust and satisfaction.

Future research should focus on implementing participant recommendations to enhance naturalness and personalization in synthetic voices. Comparative studies between trustworthy and untrustworthy voices will further refine our understanding of the specific voice parameters that impact trust in synthetic voices, enabling more effective design strategies for synthetic voice technologies.

In conclusion, this research provides valuable insights for designing synthetic voices that enhance trust and engagement in automated customer service systems, ultimately improving user experience and satisfaction.

ACKNOWLEDGMENTS

I would like to thank my supervisors, Dr. K.P. Truong and H. Garcia Goo, for their continued and invaluable guidance.

REFERENCES

[1] 2024. Journey US Voice - Google Cloud Text-to-Speech API. https://cloud.google.com/text-to-speech/docs/voice-types. Accessed: 2024-06-23.
[2] William Apple, Lynn A. Streeter, and Robert M. Krauss. 1979. Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology* 37, 5 (1979), 715–727. https://doi.org/10.1037/0022-3514.37.5.715
[3] Pascal Belin, Bibi Boehme, and Phil McAleer. 2017. The sound of trustworthiness: Acoustic-based modulation of perceived voice personality. *PLoS ONE* 12, 10 (2017), e0185651. https://doi.org/10.1371/journal.pone.0185651
[4] Derrick D.C Elkins, A.C. 2013. The Sound of Trust: Voice as a Measurement of Trust During Interactions with Embodied Conversational Agents. 22, 4 (2013), 897–913. https://doi.org/10.1007/s10726-012-9339-x

[5] Andrew J Flanagin and Miriam J Metzger. 2000. Perceptions of Internet information credibility. *Journalism & Mass Communication Quarterly* 77, 3 (2000), 515–540.

[6] Yuin Jeong, Juho Lee, and Younah Kang. 2019. Exploring Effects of Conversational Fillers on User Perception of Conversational Agents. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 1–6. https://doi.org/10.1145/3290607.3312913

[7] Xiaoming Jiang and Marc D. Pell. 1979. *Communicator physical attractiveness and persuasion. Journal of Personality and Social Psychology*. Vol. 88. 1387–1397 pages. https://doi.org/10.1037/0022-3514.37.8.1387

[8] Xiaoming Jiang and Marc D. Pell. 2017. The sound of confidence and doubt. *Speech Communication* 88 (2017), 106–126. https://doi.org/10.1016/j.specom.2017.01.011

[9] JMero (Järvinen). 2018. The effects of two-way communication and chat service usage on consumer attitudes in the e-commerce retailing sector. *Electron Markets* 28, 4 (2018), 205–217. https://doi.org/10.1007/s12525-017-0281-2

[10] Nowicki S Klofstad CA, Anderson RC. 2015. Perceptions of Competence, Strength, and Age Influence Voters to Select Leaders with Lower-Pitched Voices. 10, 8 (2015), 1. https://doi.org/10.1371/journal.pone.0133779

[11] Joris Köster, María Cobo Muñoz, Jorge Davo, Susanne Fuentes Bongenaar, and Emilie van Eps. 2024. AiCall: Democratizing AI for Efficient Inbound Call Management. https://www.linkedin.com/company/ai-call/. Accessed: 2024-05-13.

[12] Eun Ju Lee, Clifford Nass, and Scott Brave. 2000. Can computer-generated speech have gender? an experimental test of gender stereotype *(CHI EA '00)*. Association for Computing Machinery, New York, NY, USA, 289–290. https://doi.org/10.1145/633292.633461

[13] Daniel Z Levin and Rob Cross. 2003. The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management science* 50, 11 (2003), 1477–1490. https://www.researchgate.net/publication/220534702_The_Strength_of_Weak_Ties_You_Can_Trust_The_Mediating_Role_of_Trust_in_Effective_Knowledge_Transfer

[14] I. Lopatovska et al. 2021. Effects of Gendered Voices on Personality Perceptions of Conversational User Interfaces. *Journal of Voice Interaction Studies* (2021). https://www.speechinteraction.org/CHI2021/papers/CHI21-CUI%20workshop-Effects%20of%20Gendered%20Voices%20on%20Personality%20Perceptions%20of%20Conversational%20User%20Interfaces-final.pdf

[15] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An integrative model of organizational trust. *Academy of management review* 20, 3 (1995), 709–734. https://www.jstor.org/stable/258792

[16] Belin P McAleer P, Todorov A. 2014. How Do You Say 'Hello'? Personality Impressions from Brief Novel Voices. 9, 3 (2014), 1. https://doi.org/10.1371/journal.pone.0090779

[17] Erin Devers Carolyn Meeks. 2024. Gender, Voice, and Job Stereotypes. 69 (2024), 69–80. https://doi.org/10.1007/s12646-023-00765-z

[18] John W Mullennix, Steven E Stern, Stephen J Wilson, and Corrie lynn Dyson. 2003. Social perception of male and female computer synthesized speech. *Computers in Human Behavior* 19, 4 (2003), 407–424. https://doi.org/10.1016/S0747-5632(02)00081-X

[19] Clifford Nass and Scott Brave. 2005. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. The MIT Press.

[20] Kengo Ohta, Norihide Kitaoka, and Seiichi Nakagawa. 2014. Modeling filled pauses and silences for responses of a spoken dialogue system. *Int J Comput l* (2014), 998–4308.

[21] Pisanski K. Lachowicz-Tabaczek K. et al. Oleszkiewicz, A. 2017. Voice-based assessments of trustworthiness, competence, and warmth in blind and sighted adults. 24 (2017), 856–862. https://doi.org/10.3758/s13423-016-1146-y

[22] Daniel SERRA et al. 2009. *Gender pairing bias in trustworthiness*. Technical Report.

[23] V.K.M. Shiramizu, M. Lickiss, M. Halvey, and J.M. Jose. 2022. The role of valence, dominance, and pitch in social perceptions of artificial intelligence. *Scientific Reports* 12 (2022), 1594. https://pure.strath.ac.uk/ws/portalfiles/portal/151029846/Shiramizu_etal_SR_2022_The_role_of_valence_dominance_and_pitch_in_social_perceptions_of_artificial_intelligence.pdf

[24] Gabriel Skantze, Anna Hjalmarsson, and Catharine Oertel. 2013. Exploring the effects of gaze and pauses in situated human-robot interaction. In *Proceedings of the SIGDIAL 2013 Conference*. Association for Computational Linguistics, 163–172. https://aclanthology.org/W13-4029

[25] I. Torre and L. White. 2021. Trust in vocal human–robot interaction: Implications for robot voice design. In *Voice Attractiveness. Prosody, Phonology and Phonetics*, B. Weiss, J. Trouvain, M. Barkat-Defradas, and J. J. Ohala (Eds.). Springer, Singapore. https://doi.org/10.1007/978-981-15-6627-1_16

[26] Ilaria Torre, Laurence White, and Jeremy Goslin. 2016. Behavioural mediation of prosodic cues to implicit judgements of trustworthiness. In *Speech Prosody 2016*. ISCA.

A USE OF AI TOOLS

During the preparation of this work, the author used ChatGPT to review the grammar. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the work.

B LINKS

B.1 Voice Research Website

https://voiceresearch.onrender.com

B.2 Github Repository

https://github.com/Ss17TheBlank/voiceresearch