Journal of Computer Science

Original Research Paper

Automatic Re-Formulation of user's Irrational Behavior in Speech Recognition using Acoustic Nudging Model

[1]Lydia Kehinde Ajayi, [1]Ambrose Azeta, [1]Isaac Odun-Ayo, [2]Felix Chidozie and [3]Ajayi Peter Taiwo

[1]Department of Computer and Information Sciences, Covenant University, Ogun, Nigeria
[2]Department of Political Science and International Relations, Covenant University, Ogun, Nigeria
[3]Department of Mass Communication, Joseph Ayo Babalola University, Osun, Nigeria

Article history
Received: 22-08-2020
Revised: 08-12-2020
Accepted: 21-12-2020

Corresponding Author:
Lydia Kehinde Ajayi
Department of Computer and Information Sciences, Covenant University, Ogun, Nigeria
Email: lydia4reel@gmail.com

Abstract: In automatic speech recognition for development of automatic speech recognition applications, there has been numerous claims on the presence of speech recognition errors known as classified into lexical and acoustic errors. These errors distort speech signals thereby depreciating the accuracy and performance rate of speech recognition applications. Even though lexical speech recognition error problem has been partially combated, acoustic speech recognition error referred to as user's acoustic irrational behavior is being ignored causing high error rate with low accuracy which is the bone of contention and an impediment factor in the wide adoption of speech recognition technology. Users do not always behave in a rational manner especially when dealing with a particular speech recognition application. The persistent presence of these user's acoustic irrational behavior in speech have intensified the essential need to automatically detect and correct such errors, as current researches only focus on detecting user's acoustic irrational behavior but not correcting/reformulating/re-sizing this error. Hence, this paper provides an acoustic nudging model that will perform automatic correction/reformulation of user's acoustic irrational behavior in speech to achieve higher performance and accuracy using different acoustic parameters which are based in Pitch, Time gaps between words, Timbre descend and ascend time and Loudness. This study was able to discover a foundation for reducing error rate and achieve higher performance, as well as improve accuracy in speech recognition applications through detection and re-formulation of user's acoustic irrational behavior in speech signal automatically, thereby making the model applicable to any speech recognition applications. The outcome of this study would be useful in enhancing accuracy and performance in the context of automatic speech recognition.

Keywords: Acoustic Nudging Model, Automatic Speech Recognition Error, User's Acoustic Irrational Behavior, Automatic Speech Recognition, Acoustic Model

Introduction

Speech variations are either intrinsic or extrinsic variations causing Automatic Speech Recognition (ASR) error (Benzeghiba *et al*., 2006). Extrinsic variabilities occurs due to the influence from environment known as noise and intrinsic variabilities which occurrence is related to speaker's information such as age, gender, identity, health, emotional state, etc. In speech recognition system, many of the state-of-the-art speech recognition systems designed cannot match the performance of humans as they recognize human speech input but with some constraints like speaker dependency, speaker independency, speaker style and applicability to a particular task or environment (Thangarajan, 2012; Ajayi *et al*., 2020). Acoustic models may not be a good

Science Publications

© 2020 Lydia Kehinde Ajayi, Ambrose Azeta, Isaac Odun-Ayo, Felix Chidozie and Ajayi Peter Taiwo. This open access article is distributed under a Creative Commons Attribution (CC-BY) 3.0 license.

representative of speakers due to the aforementioned variations. Therefore, the question arises as to, what happens if a speaker has sore throat or stressed. Variations embedded in speech also extends beyond the phonological alterations where there can be disfluencies, false starts, repetitions, filled pauses, hesitations, etc. (Benzeghiba *et al*., 2006). Developing speech recognition system that is robust and very accurate in the presence of these variation constraints like gender, speech rate, vocal effort, accents, speaker's speech context, speaker's language, speaker's style, speaker's domain and speaker's environment is essential. Therefore, the focus of this paper is detection and correction of speech variation which is intrinsic variabilities.

Context Variability

This type of variability involves words in a language which has different meanings but includes the same phonetic realization. Their utilization is dependent on the context given (Thangarajan, 2012). This also means that their acoustic prone realization is overly dependent on neighboring phones which is caused by the physiology of articulators that is involved in production of speech sounds.

Speaker Variability

The conveyance of speech signal goes beyond just linguistic information but also information about the speaker like age, gender, health, emotional state, etc. All these make up the acoustic behavior of the speaker. For every speaker, their mode of utterance is unique in a way which is dependent on different factors like age, sex, health, education, dialect, etc. and for a speech independent recognition system, all these factors are necessary to build a combined model (Thangarajan, 2012). The complexity of vocal organs shape determines the timbre of the speaker. The location for speech signal source "the larynx" conveys pitch and other important speaker characteristics.

Environmental Variability

This type of variation affects the robustness of speech recognition systems. This has always been a huge and common speech-based interfaces especially in mobile communication devices or applications. The unpredictability of the acoustic environment variability is very high and it is unaccountable during training of acoustic models (Benzeghiba *et al*., 2006). This can cause a mismatch to occur between the test speech and the trained speech samples.

Style Variability

In isolated speech recognition system, a user can pause between words whilst speaking. It is easier to detect the spoken words boundary and also decode using silence context. In a continuous speech recognition system, it is very difficult to pause between words as words spoken cannot be detected using silence context which affects the accuracy of the system (Benzeghiba *et al*., 2006). In Speech Recognition System (SRS), the higher the speaking rate, the higher the word error rate most often referred to as inaccuracy. The emphasis on this current is on context, speaker and style variabilities which are intrinsic.

Majority of researches conducted in speech variations causing ASR errors are limited to environmental variability, detection and analysis of ASR variation errors, manual correction of lexical/phonetic ASR errors and ignoring correction of acoustic errors in speech. Even though the maturity of ASR has gotten to the stage of commercial applications with integration into many applications, high error rate with low accuracy is still a contention and an impediment in the wide adoption of speech recognition technology especially in the area of large-vocabulary continuous speech recognition or multi-speaker environment as acoustic and language models are far from being perfect (Jiang *et al*., 2013; Errattahi *et al*., 2018; Tang *et al*., 2019). The persistent presence and increase of ASR errors altering speech recognition accuracy has intensified the essential need to automatically detect and correct such errors. ASR transcription error correction is very crucial and uptmost essential not only to speech recognition accuracy enhancement and word error rate reduction but avoidance of error propagation to subsequent language processing modules such as machine translation and Human-Computer Interaction (HCI). The factors that produces this errors has been aligned from studies done as poor articulation, high degree of acoustic variability resulting in abnormal and irrational user behavior. Voice changes due to aging, illness and emotional state (angry, frustrated, joyful, sadness, tiredness, laughing, pride, guilt, relief, etc.), repetition, interruptions, channel mismatch (mismatch in recording conditions between the training and the testing speech data are the main challenges of speech recognition). All these factors corrupt the original queries given by speakers which leads to ASR errors and distortions (Jiang *et al*., 2013; Errattahi *et al*., 2018; Tang *et al*., 2019). The presence of persistent ASR errors motivates the need to find alternative techniques to assist users in automatically correcting the aforementioned error transcription. Previous work done has only made attempts to qualitatively and quantitatively detect ASR errors but has not automatically correct these errors as only manual error correction for lexical error has been suggested (Schuller, 2018; Tang *et al*., 2019). The solution proposed to these aforementioned ASR problem is to build a large targeted dataset for quantifying the detected

errors and automatically re-formulate these errors (Dasgupta, 2017; Schuller, 2018; Tang *et al*., 2019). The term re-formulation in the context of this study means automatically re-adjusting and re-sizing of speaker related errors i.e., user's acoustic irrational behavior during speech communication. This is achieved through re-formulation of the speech parameters such as Pitch, Loudness, Timbre (ascend and descend time) and Time Gaps between words measured in S, seconds that makes up human acoustic behavior through Acoustic Nudging Model.

The rest of the paper is outlined and organized as follows: Section III examines related work through the survey of speech variation (automatic speech recognition errors) detection techniques. Section IV expatiates on the acoustic nudging model and the materials and methods used in this study. Section V discusses the experimental analysis, section VI describes the results and discussion and finally, section VII highlights the recommendation and future works.

Related Work

There are different plethora of speech recognition variation errors, algorithms and technologies that have been proposed by scientific scholars and communities to enhance ASR system accuracy but are not yet robust with word error rate of up to 50% under certain conditions (Errattahi *et al*., 2018). Even though their goal is to enhance ASR system, most studies focus on detection/analysis of speech recognition variation or manual correction which are not convenient.

Kwon *et al*. (2003) analyzed emotions in speech recognition which focused on different speech features like pitch, log energy, mel-band energies and Mel Frequency Cepstrum Coefficients (MFCC) which all serves as the base features and then, added velocity/acceleration to form feature streams. The extracted features analysis was performed using Quadratic Discriminant Analysis (QDA) and Support Vector Machine (SVM). The experimental results achieved showed that pitch and energy are the most important features affecting speech recognition accuracy.

Pradier (2011) provided a theoretical and empirical approach to show the possible link between emotional speech and music perception. They analyzed that emotional speech recognition are based on pitch, timbre, loudness, intensity and dynamics from seven different emotions (neutral, sad, happy, afraid, bored, angry and disgusted) using Technische Universitat Berlin (TUB) Database and Spanish Emotional Speech (SES) database. Melanie further analyzed that musical sounds are different which are based on pitch and harmony which showed that speech sounds and music sounds has little in common.

Jiang *et al*. (2013) conducted a re-formulation queries with both lexical and phonetic changes to previous queries made by users. Further evaluation was done to measure the impacts of voice input errors in voice search and the effectiveness of different re-formulation strategies on handling these errors. The study suggested that voice input errors are needed issues to be resolved in speech recognition and the possible solution is to support user's query re-formulation. These queries are only focused on lexical and phonetic queries ignoring acoustic re-formulation and does not fully replicate mobile search environment with their given operations/tasks.

Davletcharova *et al*. (2015) proposed the detection of speech acoustic (emotions) behavior where the basic nature of speech under different emotional situations using thirty Russian male and female subjects for data collection. The subjects were asked to express certain emotional behaviors (neutral, sadness, anger and joy) as their speech were recorded using a mobile phone. The experiments were conducted in an ordinary bedroom. MATLAB was used for extracting and analyzing features from the recorded speech segments and WEKA software was used in classifying the three emotions. It was then inferred that emotional state has direct influence or alter speech signals based on speech recognition accuracy, classification accuracy and standard deviation parameters.

Dasgupta (2017) presented an algorithmic approach for detection of human emotions and quantitative analysis using voice and speech processing through several attributes which are pitch, timbre, loudness and time between words. The approach is based on three different emotional states (normal, angry and panicked) using a low sample data (two speech samples). The primary focus of the approach is to detect and analyze the deviations in the attributes used from the normal emotional state using MATLAB and Wave pad which recorded different values for both normal/neutral and other two emotional states.

Tang *et al*. (2019) presented a qualitative and quantitative analysis of speech recognition errors and subsequent user behavior on entertainment systems on voice queries from real time users which shows that length of utterances and loudness are plagued with high word error rates. The proposed approach only focused on lexical quantitative re-formulations with smaller dataset and not acoustic reformulation. Majority of researches given in speech variations causing ASR errors are limited to detection and analysis of ASR errors, manual correction of lexical/phonetic ASR errors and ignoring correction of acoustic errors in speech.

Acoustic Nudging (AN) Model

Due to the aforementioned ASR error problem, it became imperative to adapt the digital nudging

concept to form the acoustic nudging model. Digital nudging is from the concept of nudge theory originally proposed in behavioral economies but it can much more widely be adapted and applied for enabling and promoting change in humans, groups, individuals and technology (Mirsch *et al*., 2017; Inam *et al*., 2017; Ubaka-Okoye *et al*., 2020).

A nudge can be illustrated as a simple intervention within the choice architecture to steer individuals by addressing specific psychological effects and overcoming them as people does not make good decisions when they are tired, hungry, inexperienced, emotional and when common sense fails (Mirsch *et al*., 2017; Ajayi *et al*., 2019; Azeta *et al*., 2019). Whenever human nature contradicts goals, a regular real time intervention is needed to bridge that gap and keep it in check. This means, when common sense fails, common sensors is needed to bridge the gaps created.

A nudge is an intervention that must be cheap and easy to avoid with examples including giving notifications to inform people of their calorie intake either high or low, nutrition labels on food, automatic pension plan enrolment with an opt-out option and trying to putt fruit at eye level to steer individuals in choosing fruit over junk food, thereby promoting good health (Mirsch *et al*., 2017; 2018; Yamanaka and Miyashita, 2013). Other types of nudge include grabbing a coffee from Starbucks where there are options of three different available sizes (Tall, Grande and Venti). This steer individuals into been nudged by utilizing the middle option "Grande" over smallest one "Venti" or the biggest one "Tall" but it's easier to choose the middle one no matter what the absolutes sizes (Korhonen 2020). All these count as a nudge but stipulating a certain diet or exercise without a given choice (opt-out option) cannot be considered a nudge.

Nudge theory enables the re-formulation, analysis, tracking, improvement, design or re-designing of people's thinking and decision-making. This nudge theory has also been extended to the digital environment to give the concept known as digital nudging as it involves utilizing user interface design elements so as to affect user's choice by guiding people's behavior in digital choice environments through the use of user-interface design such as web-based forms and Enterprise Resource Planning (ERP) screens (Weinmann *et al*., 2016; Kroll and Stieglitz, 2019). Nudging works because people do not always behave rationally especially when dealing with a particular application. Human behavior is rational which influences their decision-making. Nudges work in digital environment by countering or altering the choice environment to change people's behavior by either giving incentives, providing feedbacks or setting defaults/threshold (Schneider *et al*., 2018).

Following the concepts of digital nudging, the proposed acoustic nudging model is built on the concept of improved digital nudging **(**Hummel *et al*., 2018**)** which involves tracking/monitoring technology in real time to monitor/track user's acoustic behavior. This theory can be applied and utilized in speech recognition as speech recognizer recognizes human speech and in doing so, the choice architecture is intervened by pulling the attention of the speech recognizer which has a detector to detect the irrational behavior features embedded in the human speech. This accentuation may trigger an automatic re-formulation which was not originally planned by the speech recognizer. This irrational behavior embedded in human speech for this study is based on five parameters which are Pitch (either low or high pitch) measured in Hz, Loudness (sound pressure level) measured in dB, Timbre (ascend and descend time) measured in S, seconds and Time between words measured in S, seconds that is embedded in each speech samples during speech generation which are considered as a significant factor that causes ASR error.

The effect of a distorted speech sample can be mitigated out to get a good sample. This step helps in correcting ASR errors based on user's behavior for both the collected speech samples (training/testing) and for any incoming speech input. The user's speech acoustic signals are re-formulated to preserve the acoustic model effectively. This step involves designing a speech sample that is not influenced by external conditions/speaking variability (user's irrational behavior) when it comes to speech recognition accuracy. It is achieved by re-formulating the speech parameters such as Pitch, Loudness, Timbre (ascend and descend time) and Time Gaps between words measured in S, seconds that makes up human acoustic behavior. The Acoustic Nudging (AN) Model is used to correct ASR errors (user's irrational behavior) in order to enhance speech recognition accuracy and reduce error rate. The system development life cycle including analysis, design, implementation and testing phase shown in Fig. 1.

The Analysis Phase

The analysis phase for the acoustic nudging model consists of four '4" requirements which sets the basis for the subsequent design phase. This phase consists of different tasks which is related to:

- Define goals to be achieved with acoustic nudging: The goal to be achieved with acoustic nudging includes detecting and correcting ASR error associated with user' irrational acoustic behavior as user' irrational acoustic behavior distort acoustic characteristics leading to low accuracy/performance which includes: Poor articulation, speaking rate variability (voice changes due to aging, illness, emotional state which can be broken down into

angry, frustrated, joyful, sadness, tiredness, laughing, pride, guilt, relief, etc.), high degree of acoustic variability (abnormal user behavior), interruptions, channel mismatch (mismatch in recording conditions between the training and the testing speech data which is the main challenge of speech recognition)

- Define and analyze how the user's behavior should be in in light of the goals to be achieved: Requirement 1 in this phase determines how the choice which is the user's behavior. This is a continuous choice which involves automatic re-formulation in order to alter or nudge off user's acoustic irrational behavior affecting speech recognition performance and accuracy. This will be achieved through tracking and altering (automatic re-formulation) the user's irrational behavior for the speech samples (training, validation and testing) and at the same time, real time automatic re-formulation for incoming speech signals without removing important contents and at the same time, making recognition faster
- Analyze user's characteristics and impediments to performing desired behavior, focusing on heuristic and biases: Heuristics can be defined as simple rules of judgements for information processing to help in surrogating complex decision making problems with easier ones (Lembcke *et al*., 2019). For this study, the heuristics to be considered based on aforementioned user's irrational behavior are Pitch, Loudness, Timbre (ascend and descend time) and Time between words measured in S, seconds (Dasgupta, 2017). Conversely, heuristics can influence the accuracy of speech recognition negatively by introducing biases (ASR error). Understanding the heuristics and biases and the potential effects of acoustic nudges can help in automatically correcting ASR errors
- Using tracking/monitoring technology in real time to monitor/track user's acoustic irrational behavior

Analyze the strengths/weaknesses of available technology channels and choose the optimal best to carry out the intervention: The appropriate channel to carry out this intervention which is the acoustic nudging is done with the aid of tensor flow application through the speech recognition application.

The Design Phase

The design phase for the acoustic nudging model consists of two '2" requirements which sets the basis for the subsequent implementation phase. Different tasks as follows:

- Select appropriate heuristics and biases (nudges) to alter user's behavior: This step includes selecting appropriate nudging mechanism to guide the speech recognizer in reformulating user's acoustics irrational behavior. Schneider *et al*. (2018) defined common nudging framework by types of choices and heuristics/bias which are broken down into binary choice (Status Quo bias known as defaults), discrete choice (Status Quo bias known as defaults, decoy effect, primary/recency effect or middle-bias options), continuous choice (anchoring/adjustment, Status Quo bias known as defaults) and any type of choice (Norms or loss aversion). For this study, continuous choice is to be utilized with heuristic/biases "anchoring and adjustment" using nudging mechanism "variation of slider endpoints" which serves as implicit anchors
- Design an intervention (acoustic nudges) to induce the desired behavior based on selected design principles: The design of the intervention (acoustic nudges) is summarized in Table 1

From Table 1, the acoustic variation slider endpoints given by the statistical analysis for detecting user's acoustic irrational behavior developed was applied and adopted in this study and at the same time, a neutral/normal speech samples from different individuals void of user's acoustic irrational behavior state (angry, frustrated, sadness, shouting, panicked, etc.) was also collated and adopted with their variation of slider endpoints values for the aforementioned heuristics and biases. The variation of slider endpoints was used in making automatic dynamic re-formulation in real time to correct ASR errors (user's acoustic irrational behavior).

Table 1: Identification of heuristics and biases with the appropriate acoustic nudging mechanism

Heuristics and biases	Acoustic nudging mechanism (Variation of slider endpoints)
Pitch	1248Hz-1355 Hz
Loudness (sound pressure level)	Gain of (-50 - 48 dB)
Timbre (ascend time)	0.12-0.06 s
Timbre (descend time)	0.11-0.05 s
Time gaps between words	0.10-0.12s

The Implementation Phase

The implementation phase for the acoustic nudging model consists of one '1" requirements which sets the basis for the subsequent testing phase. This phase task is related to the following:

- Implementation of the intervention (choice architecture) in the defined technology channel: This step includes implementing the afore-mentioned acoustic nudging choice architecture defined in the technology channel (tensor flow). The tendency term of the form is represented by Equation 1:

$$\left(Xm - Xp(5)\right) \tag{1}$$

which is added to the progonistic equation of the variable *X* where *X* represents user's rational acoustic behaviour (ASR corrected error) a substitute for user's irrational acoustic behaviour. Subscript *M* indicates the acoustic nudging model predicted value, Subscript *P*(5) indicates the acoustic nudging model prescribed value which comes from automatic re-formulation of the user's irrational acoustic behaviour after tracking context related to pitch, loudness, timbre ascend time, timbre descend time and time between gaps. The user's irrational acoustic behaviour was nudged based on the given scaling factor in Table 1. The acoustic nudging prescribed value in Table 1 is used to update the acoustic nudging model state variables after automatic re-formulation. Equation 2 is then replaced by:

$$-\left(X'm - X'p(5)\right) \tag{2}$$

where, *X*′ denotes the user's irrational acoustic behaviour (ASR error) of *X* with respect to its mean $\bar{X}$ i.e.:

$$X'm = Xm - \bar{X}m \tag{3}$$

$$X'p(5) = Xp(5) - \bar{X}p(5) \tag{4}$$

The motivation for the acoustic nudging for the user's irrational acoustic behaviour (ASR error) is that the original formula in Equation 1 can be expressed as:

$$-\left(Xm - Xp(5)\right) = \left(\bar{X}m + X'm\right) - \left(C + X'p(5)\right) \tag{5}$$

$$= \left(X'm - X'p(5)\right) - \left(\bar{X}m - \bar{X}p(5)\right) \tag{6}$$

When the model fields (heuristics and biases) are nudges towards automatic re-formulation, the first term on the right hand-side of Equation 5 can be interpreted as a forcing term tracks, detects and corrects the user's irrational acoustic behavior towards observed episodes. This is the actual purpose of using acoustic nudging in speech recognition for enhanced accuracy and performance. The 2nd term which is Equation 6 forces the acoustic nudging model mean state towards the observed mean, thereby correcting the biases in the user's speech data. The acoustic nudging model can be re-written as Equation 7:

$$-\left(X'm - X'p(5)\right) = -\left(Xm - Xp(5)\right) \tag{7}$$

where:

$$Xp(5) = Xp(5) - \bar{X}p(5) + \bar{X}m \tag{8}$$

This means that acoustic nudging model can be implemented using a term that appears identical to Equation 1 but with *Xp*(5) automatically replaced by *Xṕ* (5). It therefore requires an automatic re-formulation of the user's irrational acoustic behavior (ASR error) embedded without distorting the user's speech data.

The implementation phase for the acoustic nudging model consists of one '1" requirements which sets the basis for the subsequent testing phase. This phase task is related to the following:

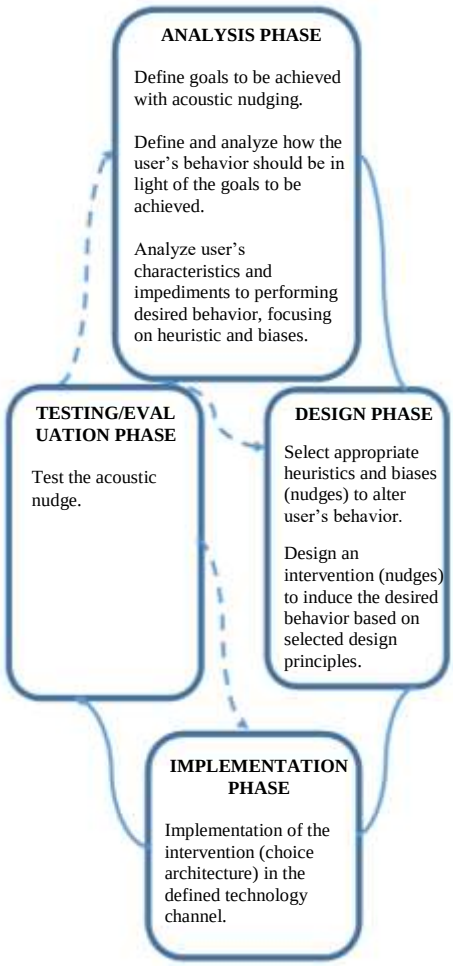

Fig. 1: Acoustic nudging design science approach

The Evaluation/Testing Phase

- Test the acoustic nudge: This step is essential to test the effect of the acoustic nudging model which is done for real-time incoming speech signal on the speech application and collected speech samples. Thorough testing is needed to get the appropriate best nudge that works best for accurate speech recognition

It is important to emphasize that all advances included in the AN model were added individually, to ensure that each advance made a difference in the process of automatic re-formulation and are working jointly to obtain better results.

Pseudocode for the Acoustic Nudging (AN) Model

The acoustic nudging algorithm (Fig. 2) is a re-formulation algorithm for the user's acoustic irrational behavior to detect and correct ASR error.

Begin
Generate the user's corrected acoustic rational behavior – (*Xm-Xp*(5))
***Input*:** Five Heuristics and biases P_1, P_2, P_3, P_4 and P_5 randomly each having different *Q* variation values.
For $e \geq 1$ **do**
  for *i* =1 to number of experiment
  el **do**
  Evaluate the desired value ($P_{(5)}$) of experiment *i*
end for
for*P* = 1 to number of variation sliderpoints Q **do**
  Q_1 = P, Q_2 = L, Q_3 = At, Q_4 = Dt and Q_5 = T_bW
  Evaluate effects of the heuristics and biases P_1, P_2, P_3, P_4 and P_5
  for P Values= 1248Hz ≤ 1355Hz ***do***
  for L= Gain of -50dB ≤ 48dB ***do***
  for At = 0.12s≤ -0.06s ***do***
  for Dt = 0.11s ≤ -0.05s ***do***
  for T_bW = Gain of 0.12s ≤ 0.10s ***do***
  for At = 0.12s ≤ -0.06s ***do***
end for
  Formulate: draw (–(*Xm-Xp*(5))) independently for every *i = 1…….. N*
  Acoustic Nudging: Choose a set of model field (heuristics and biases) $P_{(5)}$ C [N],
then compute (–(*Xm-Xp*(5)) = ($\bar{X}m$ + *X′m*)-(C + *X′p*(5))) for every *I* $\mathcal{E}$ $P_{(5)}$, where = (*Xm-Xp*(5)) = –(*X′m-X′p*(5))-($\bar{X}m$ + $\bar{X}p(5)$)) for every *i* ∈ [N]\$P_{(5)}$
Re-formulate: draw (–(*X′m-X′p*(5))-(*Xm-Xṕ*(5))) for every (*Xṕ*(5) = *Xp*(5)- $\bar{X}p(5) + \bar{X}m$) independently for *i = 1…….. N*
end for
Generation of the Acoustic Nudging (AN) parameters (-(*Xm-Xṕ*(5))))

Fig. 2: Pseudocode of the acoustic nudging model

Summary of Variables used in the Pseudocode

X: User's acoustic rational behavior
m: Acoustic nudging model predicted values
Q: Variation values
I: Number of experiments
P(5): Acoustic nudging model prescribed values for the heuristics and biases
ṕ(5): Replaced heuristics and biases
S: User's speech signal
N: Number of speech signals
X^1: User's acoustic irrational behavior
i: Number of experiment
P: Pitch
L: Loudness or sound pressure
At: Timbre ascend Time
Dt: Timbre descend time
T_bW: Time between each words

Results

Training Dataset (Acoustic Nudging Dataset): 8 (8 speakers)

The acoustic nudging modeling technique was applied on 8 speech samples from the training dataset using 8 speakers which comprises of two "2" male adult, two "2" female adult, two "2" male child and two "2" female child. They each recorded their voice-based on rational acoustic behavior in a neutral environment so as to obtain a normal/neutral values which is referred to as the acoustic rational behavior shown in Table 2 and Fig 3.

Figures 4 and 5 shows the acoustic nudging modeling process applied automatically in correcting/re-formulating a female adult's acoustic irrational test speech signals. The first chart in Fig. 4 and 5 gives an acoustic irrational behavior present in the female's speech signal and the second chart gives the automatic re-formulation of the acoustic irrational test speech signals in real time after acoustic nudging model has been applied. Figure 6 shows the acoustic modeling process applied automatically in correcting/re-formulating a male adult's acoustic irrational test speech signals. The first chart in Fig. 6 gives an acoustic irrational behavior and the second chart gives the automatic re-formulation of the acoustic irrational test speech signals.

Table 3 and 4 gives a sample value data of female and male's adult acoustic irrational behavior and the re-formulated (corrected) speech signal.

Table 2: User's acoustic rational behaviour speech samples

	Pitch (Hz)	Loudness (dB)	Timbre ascend time (s)	Timbre descend time (s)	Time gaps between words (s)
Speech sample 1(female adult)	1256	-49	0.11	0.10	0.12
Speech sample 2(male adult)	1355	-50	0.05	0.04	0.12
Speech sample 3(male child)	1248	-47	0.08	0.06	0.11
Speech sample 4(female adult)	1282	-50	0.12	0.10	0.12
Speech sample 5(female child)	1250	-47	0.11	0.10	0.10
Speech sample 6 (male child)	1324	-48	0.06	0.05	0.11
Speech sample 7(male adult)	1355	-49	0.12	0.11	0.11
Speech sample 8 (female adult)	1262	-48	0.07	0.06	0.12

Table 3: A sample data of a female and male's adult acoustic irrational behaviour

	Pitch (Hz)	Loudness (dB)	Timbre ascend time (s)	Timbre descend time (s)	Time gaps between words (s)
Speech sample 1(Tired female adult)	1244	-28	0.16	0.14	0.18
Speech sample 2(Angry male adult)	1435	-54	0.02	0.01	0.05
Speech sample 3 (Sore throat female adult)	1228	-37	0.15	0.13	0.17
Speech sample 4 (Frustrated male adult)	1356	-52	0.13	0.11	0.13
Speech sample 5 (Laughing male child)	1451	-37	0.17	0.15	0.18
Speech sample 6 (Tired female child)	1045	-41	0.04	0.03	0.06

Table 4: A sample data of an automatic re-formulated female and male's adult acoustic irrational behavior

	Pitch (Hz)	Loudness (dB)	Timbre ascend time (s)	Timbre descend time (s)	Time gaps between words (s)
Speech sample 1(Tired female adult)	1250	-49	0.12	0.11	0.11
Speech sample 2(Angry male adult)	1355	-50	0.06	0.05	0.12
Speech sample 3 (Sore throat female adult)	1260	-49	0.10	0.09	0.10
Speech sample 4 (Frustrated male adult)	1351	-50	0.12	0.10	0.10
Speech sample 5 (Laughing male child)	1312	-49	0.11	0.09	0.12
Speech sample 6 (Tired female child)	1267	-48	0.04	0.03	0.12

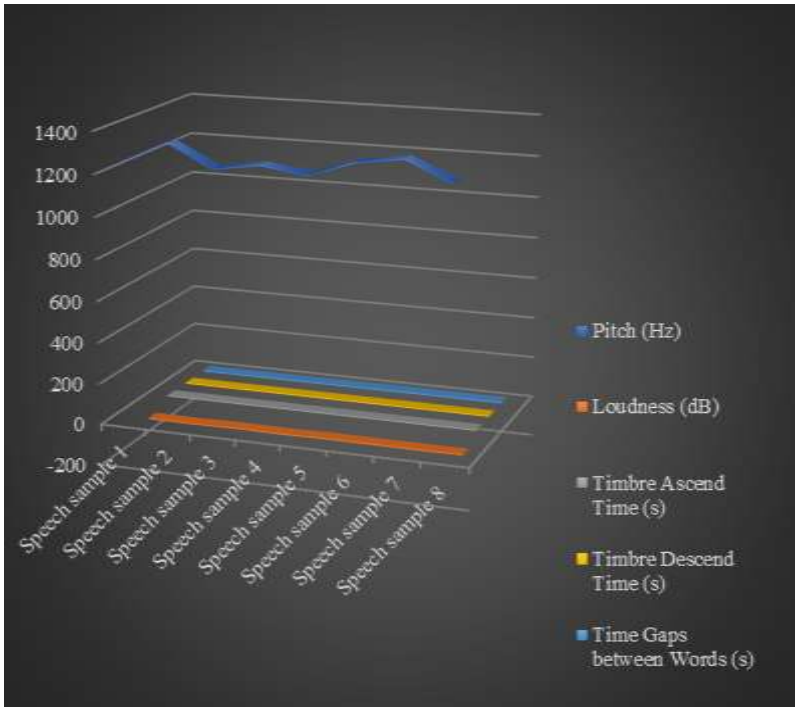

Fig. 3: User's acoustic rational (neutral) behavior speech samples

DOI: 10.3844/jcssp.2020.1731.1741

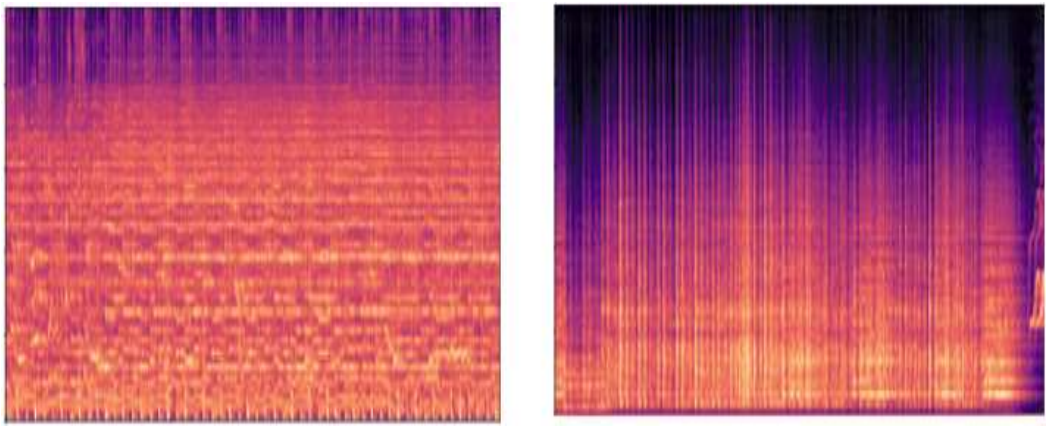

Fig. 4: Female adult's acoustic irrational and re-formulated test speech signals

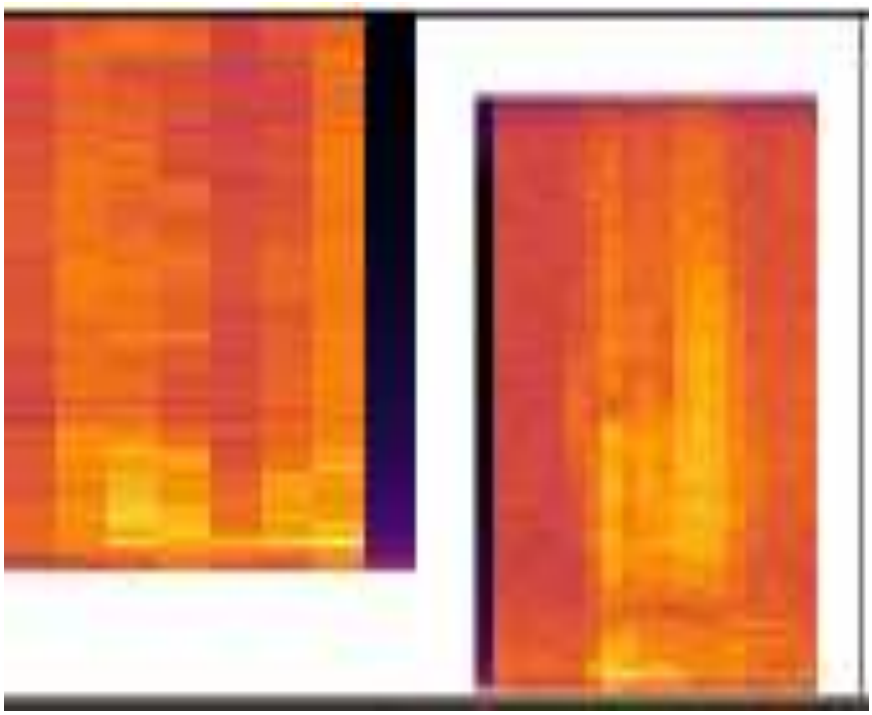

Fig. 5: Female adult's acoustic irrational and re-formulated test speech signals

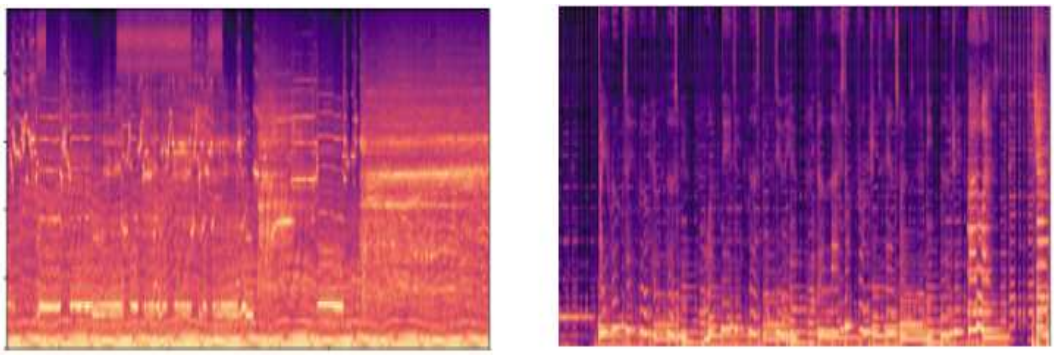

Fig. 6: Male adult's acoustic irrational and re-formulated test speech signals

Conclusion

As was presented through analyzing the tests, it is obvious that Acoustic Nudging Model is a satisfactory way to automatically re-formulate user's acoustic irrational behavior in order to achieve a higher accuracy and performance with low error rate as, it has efficiently and evidently deal with the acoustic irrational behavior (ASR error) produced by users through automatic re-adjustment and re-sizing, which is ultimately required for any speech recognition applications. This approach will help in enhancing all automatic speech recognition application performance and accuracy in the presence of any acoustic errors. For future research, implementing the concept of acoustic nudging model should be done on a real-life speech recognition application especially for continuous speech recognition application.

Acknowledgement

In this study, the researchers would like to express our deep appreciation for the support rendered by Covenant University Centre for Research, Innovation and Discovery (CUCRID).

Author's Contributions

Lydia Kehinde Ajayi: Researching and conducting the experiment as well as writing the manuscript.

Ambrose A. Azeta: Provide publication recommendations, providing guidance during the project experimentation phase, reviewing the manuscript as well as supporting the publication of the manuscript.

Isaac Odun-Ayo: Providing guidance during project experimentation phase and contributing on reviewing the manuscript.

Felix Chidozie: Did the final language-editorial work, correcting spelling errors and ensuring grammar compliance.

Ajayi Peter Taiwo: Revision of Manuscript and ensuring grammar compliance.

Ethics

This is an original manuscript which contains unpublished material. All corresponding authors have read, reviewed, provided guidance in every aspect and approved the manuscript. There are no ethical issues involved.

References

Ajayi, L. K., Azeta, A. A., Odun-Ayo, I. A., Chidozie, F. C., & Azeta, A. E. (2020). Systematic review on speech recognition tools and techniques needed for speech application development. Int. J. Sci. Technol. Res. 9, 6997-7007.

Ajayi, L. K., Azeta, A. A., Owolabi, I. T., Damilola, O. O., Chidozie, F., Azeta, A. E., & Amosu, O. (2019, August). Current Trends in Workflow Mining. In Journal of Physics: Conference Series (Vol. 1299, No. 1, p. 012036). IOP Publishing.

Azeta, A. A., Misra, S., Azeta, V. I., & Osamor, V. C. (2019). Determining suitability of speech-enabled examination result management system. Wireless Networks, 25(6), 3657-3664.

Benzeghiba, M., De Mori, R., Deroo, O., Dupont, S., Jouvet, D., Fissore, L., ... & Tyagi, V. (2006, June). Impact of variabilities on speech recognition. In Proc. SPECOM (pp. 3-16).

Dasgupta, P. B. (2017). Detection and analysis of human emotions through voice and speech pattern processing. arXiv preprint arXiv:1710.10198.

Davletcharova, A., Sugathan, S., Abraham, B., & James, A. P. (2015). Detection and analysis of emotion from speech signals. arXiv preprint arXiv:1506.06832.

Errattahi, R., El Hannani, A., & Ouahmane, H. (2018). Automatic speech recognition errors detection and correction: A review. Procedia Computer Science, 128, 32-37.

Hummel, D., Toreini, P., & Maedche, A. (2018). Improving digital nudging using attentive user interfaces: Theory development and experiment design. DESRIST 2018 Proceedings, 1-37.

Inam, I. A., Azeta, A. A., & Daramola, O. (2017, March). Comparative analysis and review of interactive voice response systems. In 2017 Conference on Information Communication Technology and Society (ICTAS) (pp. 1-6). IEEE.

Jiang, J., Jeng, W., & He, D. (2013, July). How do users respond to voice input errors? Lexical and phonetic query reformulation in voice search. In Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval (pp. 143-152).

Korhonen, M. (2020). Personality and the effectivity of digital nudges: an empirical study.

Kroll, T., & Stieglitz, S. (2019). Digital nudging and privacy: improving decisions about self-disclosure in social networks. Behaviour & Information Technology, 1-19.

Kwon, O. W., Chan, K., Hao, J., & Lee, T. W. (2003). Emotion recognition by speech signals. In Eighth European Conference on Speech Communication and Technology.

Lembcke, T. B., Engelbrecht, N., Brendel, A. B., Herrenkind, B., & Kolbe, L. M. (2019). Towards a Unified Understanding of Digital Nudging by Addressing its Analog Roots. In PACIS (p. 123).

Mirsch, T., Lehrer, C., & Jung, R. (2017). Digital nudging: Altering user behavior in digital environments. Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017), 634-648.

Mirsch, T., Lehrer, C., & Jung, R. (2018). Making digital nudging applicable: the digital nudge design method. In Proceedings of the 39th International Conference on Information Systems (ICIS). Association for Information Systems. AIS Electronic Library (AISeL).

Pradier, M. F. (2011). Emotion recognition from speech signals and perception of music. Universität Stuttgart Institut für Systemtheorie und Bildschirmtechnik Lehrstuhl für Systemtheorie und Signal verarbeitung Professor Dr.-Ing. B. Yang.

Schneider, C., Weinmann, M., & Vom Brocke, J. (2018). Digital nudging: guiding online user choices through interface design. Communications of the ACM, 61(7), 67-73.

Schuller, B. W. (2018). Speech emotion recognition: Two decades in a nutshell, benchmarks and ongoing trends. Communications of the ACM, 61(5), 90-99.

Tang, R., Ture, F., & Lin, J. (2019, July). Yelling at Your TV: An Analysis of Speech Recognition Errors and Subsequent User Behavior on Entertainment Systems. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 853-856).

Thangarajan, R. (2012). Speech Recognition for agglutinative languages. Modern Speech Recognition Approaches with Case Studies, 37-56.

Ubaka-Okoye, M. N., Azeta, A. A., Oni, A. A., Okagbue, H. I., Nicholas-Omoregbe, O. S., & Chidozie, F. (2020). Blockchain Framework for Securing E-Learning System. Institutions, 27, 28.

Weinmann, M., Schneider, C., & Vom Brocke, J. (2016). Digital nudging. Business & Information Systems Engineering, 58(6), 433-436.

Yamanaka, S., & Miyashita, H. (2013, October). The nudging technique: input method without fine-grained pointing by pushing a segment. In Proceedings of the adjunct publication of the 26th annual ACM symposium on User interface software and technology (pp. 3-4).