Journal of Robotics and Control (JRC)
Volume 1, Issue 4, July 2020
ISSN: 2715-5072, DOI: 10.18196/jrc.1423

SMART-In English: Learn English Using Speech Recognition

Dhanar Intan Surya Saputra[1], Sitaresmi Wahyu Handani[2], Kuat Indartono[3], Andik Wijanarko[4]
[1, 2, 3, 4] Department of Informatics, Universitas Amikom Purwokerto, Central Java, Indonesia
Email: dhanarsaputra@amikompurwokerto.ac.id

***Abstract*— English is an international language and important to learn. For someone learning English sometimes is a difficulty, especially in pronunciation. Therefore, SMART-In is a prototype of Android App that uses Speech Recognition technology by utilizing services from the Cloud Speech API (Application Programming Interface). SMART-In English can be used as an alternative to English learning, especially in the pronunciation of a word. Using speech recognition can display the score of the pronunciation spoken by the user, recorded, show a level the pronunciation of the word and display the correct pronunciation.**

Keywords*— *SMART-In English, Pronunciation, Speech Recognition, Cloud Speech.

I. INTRODUCTION

For some people will find it difficult to Learning vocabulary pronunciation in English that is not an original language. Differences in the habit of speaking are one of the causes of frequent difficulties in English pronunciation.

In Bahasa has two systems: vowel and consonant whereas in English has more than 2 vocal systems, diphthongs, consonants and clusters. If in Bahasa only recognize the vowels / a / - / i / - / u / - / e / - / o / but in English many vowel / - / ɑ: / - / ɒ / - / ɜ: / - / i: / - / ʊ / - / u: / - / ʌ / - / I / and the vowel is not easy to pronounce for some people [1]–[5].

The development of technology, can be applied in various aspects of life, such as creating tools, use in industry [6], [7], to creating renewable energy [8] to help create learning media. Learning Method of pronunciation is increasing and growing, in this study build a prototype SMART-In English, as a medium of learning using Android-based mobile application. The prototype is built using speech recognition technology that utilizes Cloud Speech API (Application Programming Interface).

Google as the largest and number one search engine in the world and continues to evolve its services. Cloud Speech API is part of the Google Cloud Platform, a service that consists of the main components to build cloud-based applications. These services are Google AppEngine, Google Compute Engine, Google Cloud Storage and Google BigQuery, all these services are intended for developers who want to integrate Google services into their apps [9] [10][11][12].

In Pre-trained Machine Learning Models, Google Translate API and Cloud Vision API, have been integrated into the Google Cloud Speech API. With such a complete API, developers can develop applications that can view, hear, and translate [13].

The Cloud API determines how apps communicate with cloud computing. Cloud API offers a way the apps can request information from the platform and use it's the facilities [14]. Cloud Computing is the use of computer technology and the development of Internet-based services that allow accessing on demand to large collection of resources that can be set quickly with minimal management efforts [15][16].

The utilization of the cloud API makes it possible to apply in education and teaching and learning activities [10]. For example, the use of online document editors using the Web API [17]. It is useful for the development of learning activities and the development of knowledge to students as well as increasing motivation [18]. Some of the benefits of the Google Cloud API service in education, which provides motivation, fun, capabilities in information technology, ease of use, cost savings, protected privacy, guaranteed security and new breakthroughs or innovation [19].

Google Cloud Speech API can be used as media development Learn English pronunciation vocabulary using speech recognition technology. Speech Recognition is the process of converting voice signals into machine language in the form of digital data (usually simple text) [20] [21], is an interdisciplinary subfield of computational linguistics that develops technologies and methodologies that enables the recognition and translation of spoken language and voice into text by computers or applications [22] [23]. Voice recognition [24]–[28] implies the ability to match patterns from acquired or acquired vocabularies to sound signals into the proper form so that computers or other machines can recognize words spoken by humans [29] [30]. This can be applied in SMART-In English, an interactive application that supports and simplifies the process of learning vocabulary pronunciation in English. Users can be more interested, application and optimize the learning process, as well as provide opportunities for users to be more confident in learning as well as broaden the horizons.

II. METHODOLOGY

This research is arranged in a systematic and phases by stages in the framework of research. Beginning by analyzing

and identifying problems, designing and making applications, implementing through evaluation (Figure 1).

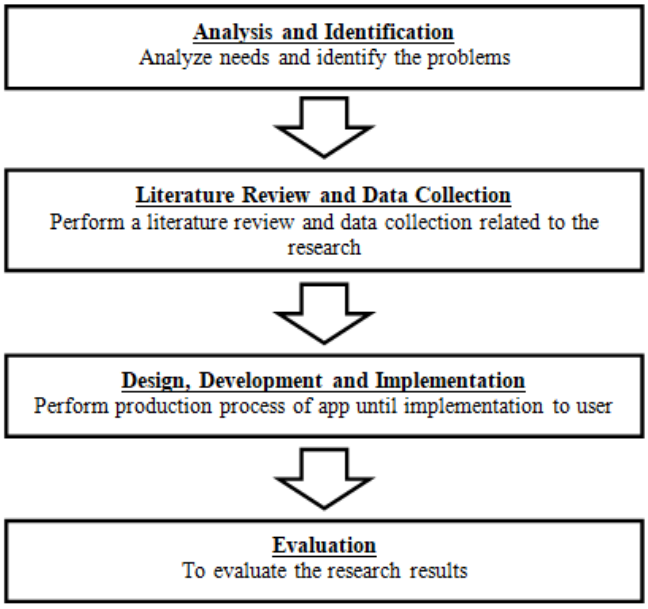

Fig. 1. Framework of Research

III. RESULTS AND DISCUSSION

Good pronunciation of English is a must, because if the vocab and grammar are correct, but the other person does not understand the message then communication will not go on. Difficulty pronouncing according to the way native speakers say a word or phrase, often making a learner feel inhibited and reducing his confidence when engaging in an English conversation, needs repetitive and frequent exercises to achieve maximum results.

From the analysis, SMART-In English is developed with the concept of a single user, as for an architecture system which can be described is as in Figure 2.

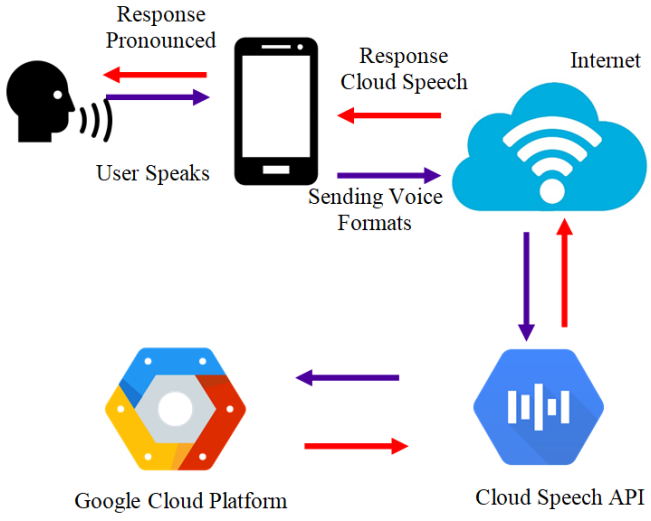

Fig. 2. Systems Architecture

Development of applications from small scale to enterprise scale from time to time will face challenges. These challenges include: challenges to improve productivity development, challenges in responding to growing demands, challenges to maintain the sustainability of the value of an ongoing information system, and challenges to maintain system security. For this reason, the role of software architecture is very important to resolve these issues.

Software architecture describes the structure of the system such as: (1) Software elements portrayed as abstractions and in the form of system modules or high-level components; (2) External Visible Properties of elements that describe the features of the elements being exposed and represent the services provided to other elements; and (3) Relationships of elements that describe the way elements interact with each other.

Figure 2 is a general system architecture, users can record the voice in English, then Android capture and transmit voice over the internet then process it via the Cloud Speech API in the Google Cloud Platform, if the sound is identified then will respond and sends back to the user, SMART-In English will display the results of the spoken pronunciation.

A. *SMART-In English Development*

Use case diagram, as illustrated in Figure 3 is a path of an app in which the player is described as the user and when the player pushes the button, then there are several "class" consisting of 4 class and when the player presses the button of about the content then display of the app.

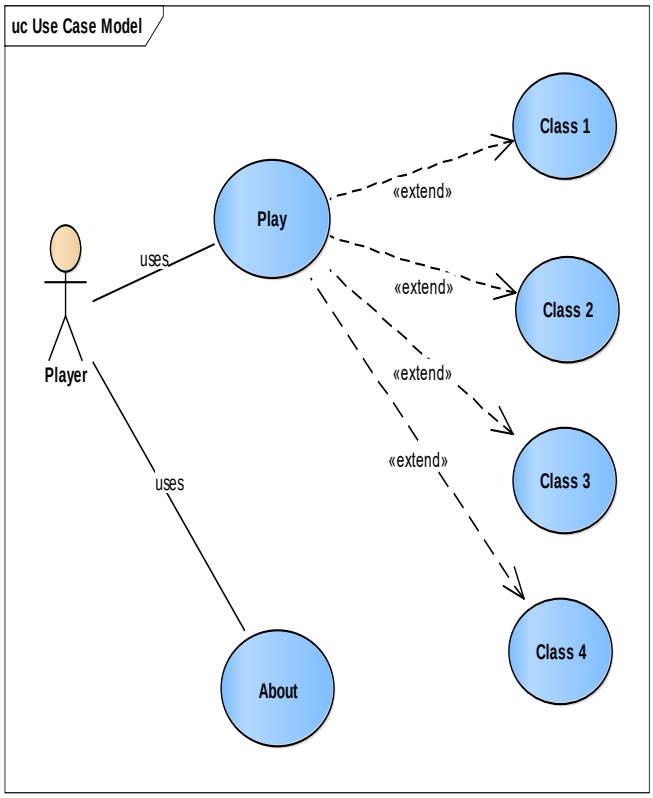

Fig. 3. Use Case Diagram SMART-In English

The sequence diagram is depicted in Figure 4. Player presses open (UI) then becomes Open (Play) then wait some time then query request back question again open the

question page display answers come in answer data request display answer return check back return the completed application request and then finish.

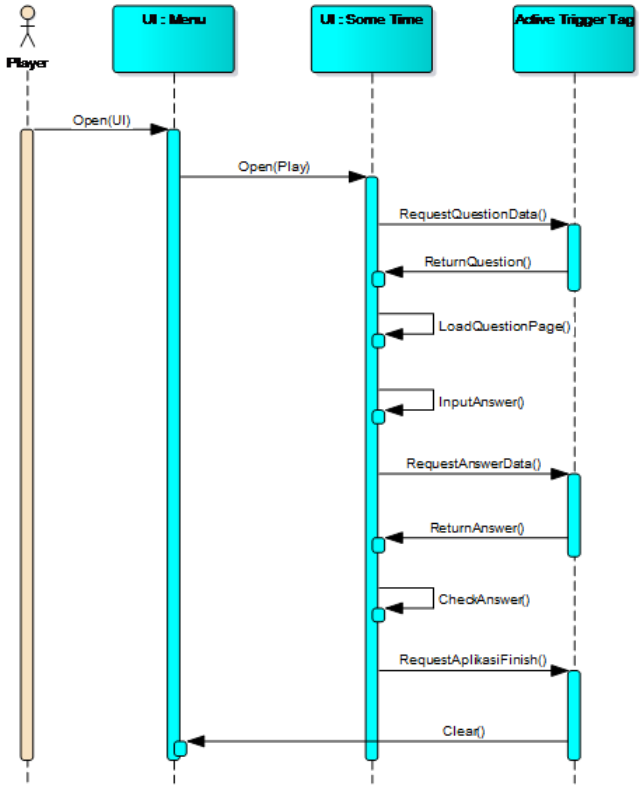

Fig. 4. Sequence Diagram SMART-In English

B. *Implementation of SMART-In English*

The results of the analysis, design, and manufacture of the application generated SMART-In English as a medium of learning English pronunciation using speech recognition that utilizes Cloud Speech API.

1) *Structure Navigation*

There are menus (Figure 5) in SMART-In English, created to meet the needs of learning English pronunciation. The menu is designed based on the object category that will be displayed on SMART-In English. The menu displays several functions that can be used by users in the application.

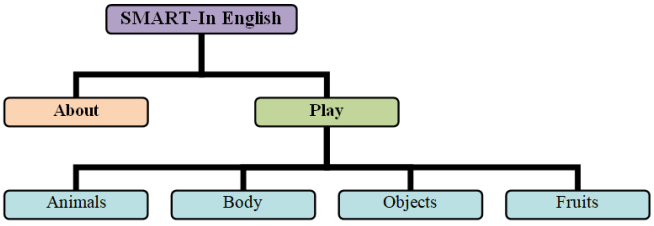

Fig. 5. Structure Navigation in SMART-In English

2) *Menus*

When the user opens the SMART-In English application, the user will go to the main page of Main Menu (Figure 6, (a)) and Play Menu (Figure 6 (b)), there is a choice of learning material, i.e. Animals, Body's, Objects and Fruits.

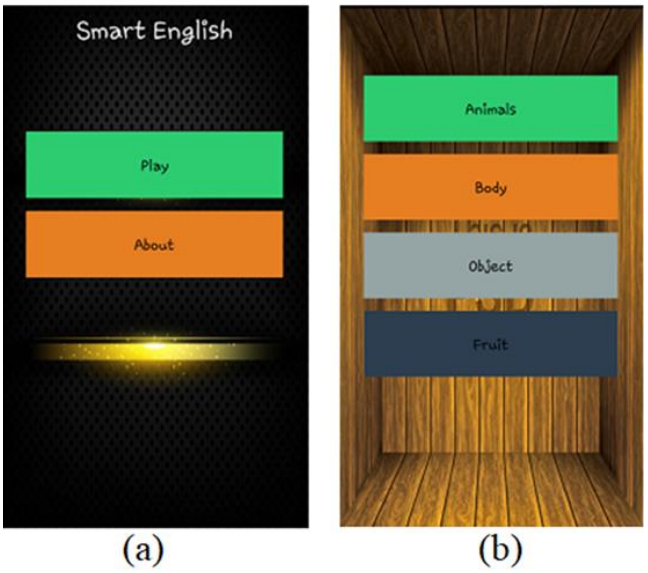

Fig. 6. Menus SMART-In English [10]

3) *Usage*

In the use of SMART-In English application users simply select the menu of Animals, Body, Objects, and Fruits. Figure 7 there are categories and displaying the correct way of pronunciation, (a) showing the way of pronunciation \ dog \; (b) shows how to pronounce \ laptop \; (c) how to pronunciation \ lemon \.

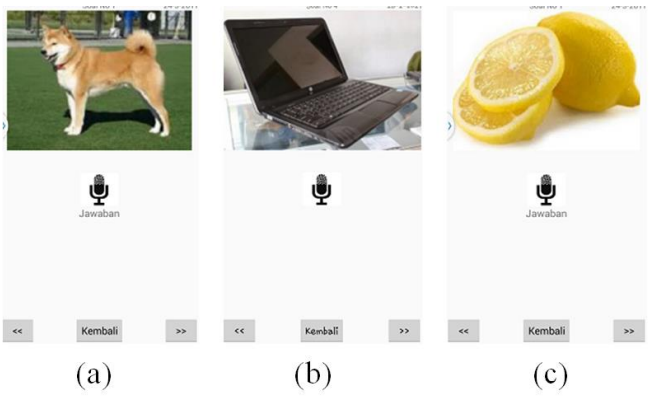

Fig. 7. Menu Pronunciation in SMART-In English

SMART-In English will show whether the pronunciation is true or false and will display the score of pronunciation being pronounced (Figure 8). Then the app will display the correct pronunciation and give the value (score) similarity if the user pronounces the word wrong (Figure 9).

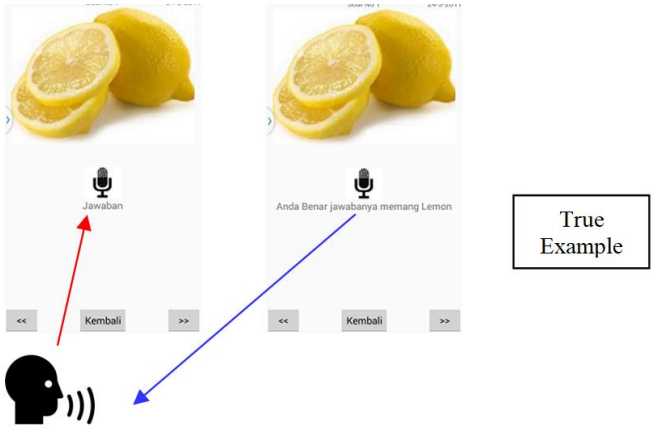

Fig. 8. Examples of Correct Pronunciation in SMART-In English

A message if the spoken word is correct, that is "*Anda Benar*"

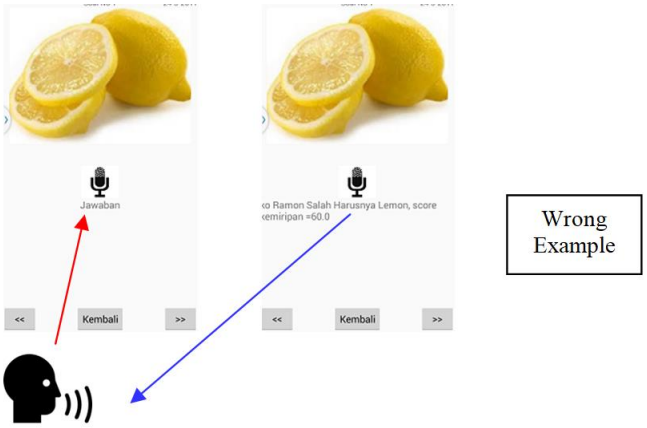

Fig. 9. Examples of Wrong Pronunciation in SMART-In English

The message if the word is spoken wrong will appear, i.e. "*Jawaban Anda Salah*"

C. *Evaluate and Testing*

The main purpose of software testing is actually simple, namely to ensure that the software produced matches the requirements that were previously determined. When the requirements of a system have been compiled then there should be a test planning. In addition, a testing process requires a final goal that can be assessed so that the tester can stop doing a test when those goals are achieved.

This stage is very important, because at this stage has the main goal is to ensure the functions of the components of the system has functioned in accordance with the expected and in accordance with the concept. There are 2 stages to do that is testing the application side and user acceptance testing of the application.

1) *Testing Application*

In this stage testing done using black box testing method. This black box method is a test program based on the function of the program. The purpose of this method of black box testing is to find function errors in the program, Table I and Table II.

TABLE I. BLACK BOX TESTING

No	Process	Expected Results	Result
1	Run App	Splash Screnn run to menus	Fit
2	Menu "Play"	Class Run	Fit
3	Menu "About"	Run About	Fit
4	Menu "Exit"	Exit App	Fit
5	Class "1"	Run Class 1	Fit
6	Class "2"	Run Class 2	Fit
7	Class "3"	Run Class 3	Fit
8	Class "4"	Run Class 4	Fit

TABLE II. BLACK BOX RESULTS

No	Procese	Expected Results	Result
1	Choice class	Run	Fit
2	Page	Questions	Fit
3	Answering	Correct Answer	Fit
4	Answering	Wrong Answer	Fit
5	Push Back	Back	Fit
6	Questions	Next Questions	Fit
7	Finish	Score	Fit

2) *User Acceptance*

From the results of acceptance testing to the user, shows the final results of testing has a performance in accordance with expectations and feasible to be used as an alternative media in learning English pronunciation.

IV. CONCLUSION AND FUTURE STUDY

SMART-In English has been successfully implemented in utilizing the cloud speech API for the development of English pronunciation learning media using Speech Recognition technology. SMART-In English is still limited to using one word in English, hoping that in subsequent research it can be better developed for word combinations or use of English sentences. In addition, the addition of vocabulary is also important to be a concern in the development of this application.

REFERENCES

[1] J. Monte-Ordoño and J. M. Toro, "Different ERP profiles for learning rules over consonants and vowels," *Neuropsychologia*, vol. 97, no. February, pp. 104–111, Mar. 2017.

[2] T. Nazzi and L. Polka, "The consonant bias in word learning is not determined by position within the word: Evidence from vowel-initial words," *J. Exp. Child Psychol.*, vol. 174, pp. 103–111, Oct. 2018.

[3] T. D. Wewalaarachchi and L. Singh, "Vowel, consonant, and tone variation exert asymmetrical effects on spoken word recognition: Evidence from 6-year-old monolingual and bilingual learners of Mandarin," *J. Exp. Child Psychol.*, vol. 189, p. 104698, Jan. 2020.

[4] Z. Shirzhiyan, E. Shamsi, A. S. Jafarpisheh, and A. H. Jafari, "Objective classification of auditory brainstem responses to consonant-vowel syllables using local discriminant bases," *Speech Commun.*, vol. 114, no. September, pp. 36–48, Nov. 2019.

[5] J. Feng *et al.*, "Effect of blindness on mismatch responses to Mandarin lexical tones, consonants, and vowels," *Hear. Res.*, vol. 371, pp. 87–97, Jan. 2019.

[6] V. M. Cáceres *et al.*, "Daily zero-reporting for suspect Ebola using short message service (SMS) in Guinea-Bissau," *Public Health*, vol. 138, pp. 69–73, Sep. 2016.

[7] S. J. Iribarren *et al.*, "Scoping review and evaluation of SMS/text

messaging platforms for mHealth projects or clinical interventions," *Int. J. Med. Inform.*, vol. 101, pp. 28–40, May 2017.
[8] S. Zhou and B. D. Solomon, "Do renewable portfolio standards in the United States stunt renewable electricity development beyond mandatory targets?," *Energy Policy*, vol. 140, no. February, p. 111377, May 2020.
[9] J. Peng, X. Zhang, Z. Lei, B. Zhang, W. Zhang, and Q. Li, "Comparison of several cloud computing platforms," in *2009 Second international symposium on information science and engineering*, 2009, pp. 23–27.
[10] D. I. S. Saputra, S. W. Handani, and G. A. Diniary, "Pemanfaatan Cloud Speech Api Untuk Pengembangan Media Pembelajaran Bahasa Inggris Menggunakan Teknologi Speech Recognition," *Telematika*, vol. 10, no. 2, pp. 92–105, 2017.
[11] S. P. T. Krishnan and J. L. U. Gonzalez, *Building Your Next Big Thing with Google Cloud Platform: A Guide for Developers and Enterprise Architects*. Springer, 2015.
[12] C. Ratcliff, "What are the top 10 most popular search engines," *Retrieved from Search Engine Watch https//searchenginewatch. com/2016/08/08/what-are-the-top-10-mostpopular-search-engines*, 2016.
[13] R. Arulmurugan, K. R. Sabarmathi, and H. Anandakumar, "Classification of sentence level sentiment analysis using cloud machine learning techniques," *Cluster Comput.*, pp. 1–11, 2017.
[14] N. Mithapelli, S. Chavan, and J. Kumari, "Alumni Tracking Using Google Map API and Social Media based on GPS and LBS," *Int. J. Eng. Sci.*, vol. 25, no. 11, 2016.
[15] D. Petcu, C. Craciun, and M. Rak, "Towards a cross platform cloud API," in *1st International Conference on Cloud Computing and Services Science*, 2011, pp. 166–169.
[16] D. I. S. Saputra, E. Utami, and A. Sunyoto, "Penerapan Mobile Augmented Reality Berbasis Cloud Computing Pada Harian Umum Radar Banyumas," in *Seminar Nasional Informatika (SEMNASIF) 2015*, 2015.
[17] R. H. Rizzardini, C. Gütl, and H. R. Amado-Salvatierra, "Using Cloud-Based Applications for Education, a Technical Interoperability Exploration for Online Document Editors," in *International Workshop on Learning Technology for Education in Cloud*, 2015, pp. 219–231.
[18] S. W. Handani, M. Suyanto, and A. F. Sofyan, "Penerapan konsep gamifikasi pada e-learning untuk pembelajaran animasi 3 dimensi," *Telematika*, vol. 9, no. 1, 2016.
[19] M. E. Brown and D. L. Hocutt, "Learning to use, useful for learning: a usability study of Google apps for education," *J. Usability Stud.*, vol. 10, no. 4, pp. 160–181, 2015.
[20] D. Povey *et al.*, "The Kaldi speech recognition toolkit," in *IEEE 2011 workshop on automatic speech recognition and understanding*, 2011, no. CONF.
[21] E. (Betsy) A. Baker, "Apps, iP ads, and Literacy: Examining the Feasibility of Speech Recognition in a First-Grade Classroom," *Read. Res. Q.*, vol. 52, no. 3, pp. 291–310, 2017.
[22] B. R. Reddy and E. Mahender, "Speech to text conversion using android platform," *Int. J. Eng. Res. Appl.*, vol. 3, no. 1, pp. 253–258, 2013.
[23] W. Diao, X. Liu, Z. Zhou, and K. Zhang, "Your voice assistant is mine: How to abuse speakers to steal information and control your phone," in *Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones & Mobile Devices*, 2014, pp. 63–74.
[24] Y. Xue, Y. Hamada, and M. Akagi, "Voice conversion for emotional speech: Rule-based synthesis with degree of emotion controllable in dimensional space," *Speech Commun.*, vol. 102, no. June, pp. 54–67, Sep. 2018.
[25] S. H. Mohammadi and A. Kain, "An overview of voice conversion systems," *Speech Commun.*, vol. 88, pp. 65–82, Apr. 2017.
[26] K. Kobayashi, T. Toda, and S. Nakamura, "Intra-gender statistical singing voice conversion with direct waveform modification using log-spectral differential," *Speech Commun.*, vol. 99, no. March, pp. 211–220, May 2018.
[27] N. J. Shah and H. A. Patil, "A novel approach to remove outliers for parallel voice conversion," *Comput. Speech Lang.*, vol. 58, pp. 127–152, Nov. 2019.
[28] F.-L. Xie, F. K. Soong, and H. Li, "Voice conversion with SI-DNN and KL divergence based mapping without parallel training data," *Speech Commun.*, vol. 106, no. November 2018, pp. 57–67, Jan. 2019.
[29] J. Grudin, "Human-computer interaction," *Annu. Rev. Inf. Sci. Technol.*, vol. 45, no. 1, pp. 367–430, 2011.
[30] R. Aihara, T. Nakashika, T. Takiguchi, and Y. Ariki, "Voice conversion based on non-negative matrix factorization using phoneme-categorized dictionary," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 7894–7898.