



# Voice Recognition for Periodontal Probing Medical Records under Korean–English Bilingual Conditions: A Feasibility Study

Young Woo Kim<sup>1,\*</sup>, Jin Hyeok Kook<sup>2,\*</sup>, Yiseul Choi<sup>3,4</sup>, Wonse Park<sup>3,4</sup>

<sup>1</sup>School of Mechanical Engineering, Yonsei University, Seoul, Korea

<sup>2</sup>Innovaid Co., Seoul, Korea

<sup>3</sup>Department of Advanced General Dentistry, Yonsei University College of Dentistry, Seoul, Korea

<sup>4</sup>Yonsei Institute for Digital Health, Yonsei University, Seoul, Korea

**Objectives:** This study evaluated the feasibility of voice recognition-based electronic medical record (EMR) documentation for periodontal probing in dentistry, particularly emphasizing Korean-English bilingual speech patterns and real-world clinical conditions. **Methods:** Experiments were conducted in a dental chair setting during routine clinical hours. Environmental noise levels were measured, and two microphone types (stationary and pin-type) were evaluated. Periodontal probing phrases composed of three-digit numbers and positional terms were used for speech recognition. Consistent with common clinical practice in Korea, numerical values were spoken in Korean, whereas positional terms were spoken in English. Two speech-to-text application programming interfaces, Google Cloud Speech-to-Text and Naver Clova Speech Recognition, were assessed. Recognition accuracy was evaluated for both numerical components and complete bilingual phrases. **Results:** The mean environmental noise level was 60.65 dB and was minimally influenced by activity at adjacent dental chairs. The stationary microphone failed to capture speech effectively, whereas the pin-type microphone demonstrated stable recognition performance. For three-digit number recognition, accuracy was 88.3% with Google and 96.8% with Naver. For full-phrase recognition, complete matching was achieved in 36.7% of cases for Google and 52.5% for Naver. Partial recognition occurred more frequently for numerical components than for English positional terms. **Conclusions:** Voice recognition-based EMR documentation for periodontal probing demonstrated preliminary feasibility in a dental clinical environment; however, performance was influenced by Korean-English bilingual speech patterns. These findings suggest that bilingual speech characteristics should be considered when implementing voice recognition systems in dental EMR workflows. Further optimization is required before routine clinical application.

**Keywords:** Voice Recognition, Natural Language Processing, Electronic Health Records, Dentistry, Multilingualism

**Submitted:** January 13, 2026, **Revised:** March 18, 2026, **Accepted:** April 21, 2026

## Corresponding Author

Wonse Park

Department of Advanced General Dentistry, Yonsei University College of Dentistry, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea. Tel: +82-2-2228-8980, E-mail: wonse@yuhs.ac (<https://orcid.org/0000-0002-2081-1156>)

\*These authors contributed equally to this work.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## I. Introduction

Electronic medical records (EMRs) have been widely adopted in both medical and dental practice, enabling longitudinal tracking of patient data and more comprehensive data management [1,2]. However, real-time EMR documentation during clinical procedures remains burdensome. In many settings, handwritten charting followed by later transcription is still common, leading to inefficiencies and an increased risk of documentation errors [3-5].

In dentistry, EMR documentation presents additional challenges because clinicians must maintain direct contact with the patient's oral cavity while adhering to strict infection control protocols. Input devices such as keyboards and mice are often located in areas exposed to patient-generated aerosols, necessitating frequent glove changes or the involvement of additional personnel for documentation [6]. Accordingly, hands-free EMR recording methods may improve both clinical workflow efficiency and infection control.

Another characteristic of dental EMRs is the high proportion of numerical data, including tooth position, periodontal pocket depth, and gingival margin measurements. This reliance on numerical information may increase the likelihood of documentation errors, particularly during busy clinical workflows. Voice recognition, or speech-to-text technology, has therefore been explored as a hands-free alternative for EMR documentation, with reported benefits including reduced transcription time and improved clinician satisfaction across several medical fields [4,7-12].

Despite these potential advantages, the application of voice recognition in dentistry involves unique linguistic challenges. In Korea, although Korean is the primary language, English terminology is widely used in medical documentation, whereas numerical values are typically spoken in Korean. Given the numerical intensity of dental charting, this mixed Korean-English speech pattern may adversely affect voice recognition performance.

Periodontal probing depth recording exemplifies this bilingual complexity. As illustrated in Figure 1, documentation consists of numerical tooth identifiers, English positional terms, and numerical depth values, which are typically spoken using a hybrid Korean-English format in clinical practice. The objective of this study was to evaluate the feasibility of implementing a voice recognition-based EMR documentation system for periodontal probing in dentistry, with a focus on clinical environmental conditions and Korean-English bilingual speech patterns. This study emphasizes characterization of recognition behavior under bilingual

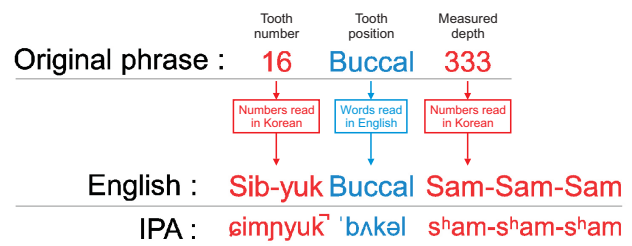


Figure 1. Example of bilingual pronunciation for periodontal probing records in Korean clinical practice. Numerical values are pronounced in Korean, whereas positional terms (e.g., buccal) are expressed in English. The figure illustrates the mixed Korean-English structure of clinical speech, along with the corresponding representations in standard English orthography and the International Phonetic Alphabet (IPA).

clinical speech rather than formal comparison of system performance.

## II. Methods

### 1. Environmental Testing

Environmental noise measurements were conducted at Yonsei University Dental Hospital in Korea. A digital sound level meter (DT-95; CEM Instruments, West Bengal, India), capable of detecting sound levels from 35 to 130 dB, was used. Measurements were obtained during routine clinical hours under two surrounding conditions: (1) adjacent dental chairs were vacant and (2) teeth-scaling procedures were being performed at nearby chairs. Of 311 total measurements, 28 values identified as high or low outliers were excluded from the analysis. Outliers were defined using the interquartile range (IQR) method; values below  $Q1 - 1.5 \times IQR$  or above  $Q3 + 1.5 \times IQR$  were classified as outliers and excluded.

Following the noise assessment, two voice-acquisition microphones were evaluated under identical clinical conditions to assess their feasibility for use in the dental environment. The devices included a stationary microphone (TITAN STUDIO PRO; XENICS Creative, Seoul, Korea) and a pin-type lapel microphone (MP15; ABKO, Seoul, Korea). The stationary microphone was positioned approximately 1 m above the ground and 1 m from the presumed noise source to minimize interference with patient and practitioner movement. The pin-type microphone was attached to the practitioner's lapel near the sound source. Technical specifications for the microphones are provided in Supplementary Table S1. Each microphone was tested five times using Korean sentences relevant to dental charting.

The experimental setting was designed to closely reflect

routine clinical practice. Practitioners wore masks, and speech was produced naturally without deliberate adjustment for microphone placement. Each test sentence was spoken once. Speech recognition was evaluated using the Google Cloud Speech-to-Text application programming interface (API) with randomly selected English and Korean texts.

## 2. Evaluation of Numbers by API Models

Periodontal probing depth documentation consists of a three-digit numerical component followed by a positional term (mesial, middle, or distal). In accordance with common clinical practice in Korea, numerical values were pronounced in Korean, whereas positional terms were pronounced in English (Figure 1).

Two speech recognition APIs were evaluated: Google Cloud Speech-to-Text and Naver Clova Speech Recognition. The purpose of this evaluation was not to perform formal benchmarking of the two models, but to examine differences in recognition characteristics under Korean-English bilingual speech conditions.

Both speech-to-text APIs were used with their default settings, without additional model training or customization. For Google Cloud Speech-to-Text, automatic language detection was enabled to accommodate mixed Korean-English input, whereas Naver Clova Speech Recognition was used in its standard Korean recognition mode.

No text normalization or post-processing (e.g., correction of spacing, punctuation, or number formatting) was applied, and raw recognition outputs were used for analysis. Numerical inputs were evaluated as three-digit sequences without digit-level correction or transformation.

Speech data were recorded using a lapel-mounted microphone positioned near the speaker under real clinical conditions, including mask wearing and ambient background noise. All utterances were captured in real time and processed directly by each API without intermediate filtering or preprocessing.

Number-only recognition tests and full-phrase recognition tests were conducted sequentially. In the number recognition tests, a result was classified as a match only when all three digits were correctly recognized. For full-phrase recognition, test phrases were randomly selected from actual periodontal probing records. Each phrase consisted of Korean-pronounced numerical values followed by an English positional term (e.g., “16 mesial,” pronounced as “sip-yook mesial”). A total of 729 phrases containing numbers were evaluated for number recognition, and 120 full phrases were evaluated for

bilingual phrase recognition.

Speech samples were obtained from two male participants involved in clinical dental practice. All participants performed the speech tasks under identical conditions, including wearing surgical masks and using a pin-type microphone in a real clinical environment. Each participant produced the same set of predefined utterances, and the distribution of utterances was kept consistent across speakers. No speaker-level stratified analysis was performed.

## III. Results

### 1. Environmental Test Results

The results of environmental noise measurements under the two clinical conditions are summarized in Figure 2. After exclusion of outliers, the mean noise level was 60.65 dB. When stratified by the number of adjacent noise sources, the mean noise level was 59.5 dB with no active neighboring chairs, 61.0 dB with one active chair, and 61.2 dB with two active chairs. Noise variability was slightly greater in the absence of neighboring noise sources (range, 54.3–64.2 dB), whereas activity at adjacent chairs did not substantially alter overall noise levels.

In microphone feasibility testing, the stationary microphone failed to recognize speech in all five attempts, whereas the pin-type microphone achieved consistent recognition. Given the close proximity of dental chairs, a distance of approximately 1 m created greater separation between the practitioner and the stationary microphone than between the microphone and surrounding noise sources. Consequently, subsequent experiments were conducted using the pin-type microphone.

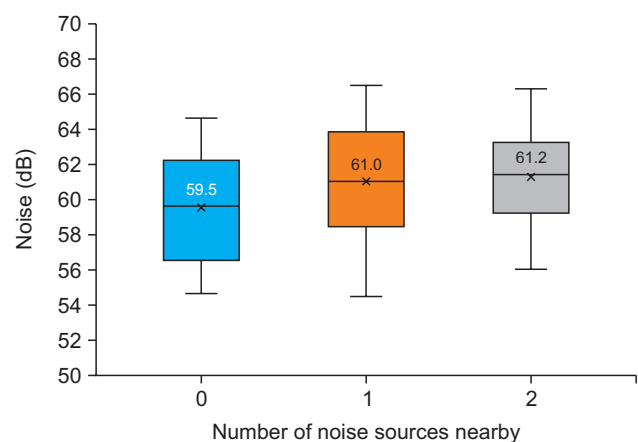


Figure 2. Results of noise measurements under different numbers of nearby noise sources.

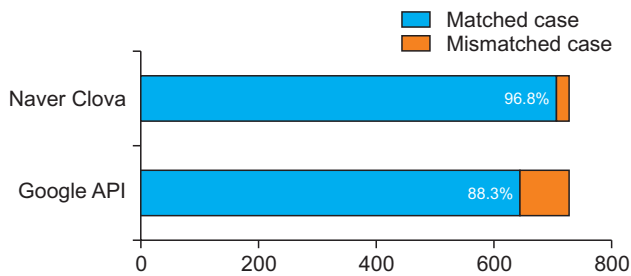


Figure 3. Results of the three-digit number recognition test.

## 2. Evaluation of Number Recognition

Three-digit number recognition accuracy for each API model is presented in Figure 3. Accuracy was 88.3% for Google Cloud Speech-to-Text and 96.8% for Naver Clova. Among mismatched cases (85 for Google Cloud Speech-to-Text and 23 for Naver Clova), nine cases were incorrectly recognized by both models. Detailed accuracy by digit position and numerical value is provided in Supplementary Table S2.

## 3. Evaluation of Korean–English Bilingual Recognition

The results of bilingual phrase recognition are summarized in Figure 4. Recognition outcomes were categorized into four groups: complete match, number-only match, word-only match, and complete mismatch. Complete phrase recognition was observed in 52.5% of cases for Naver Clova and 36.7% for Google Cloud Speech-to-text. Among partially matched cases, numerical components were recognized more frequently than English positional terms (Google 49.2% vs. 12.5%; Naver Clova 30.0% vs. 15.8%). Detailed results for each phrase are presented in Supplementary Table S3.

## IV. Discussion

The objective of this study was to evaluate the feasibility of applying voice recognition technology to dental EMR documentation under Korean-English bilingual conditions in a real clinical environment. Two speech recognition APIs, Google Cloud Speech-to-Text and Naver Clova Speech Recognition, were evaluated to characterize recognition performance under bilingual speech input. The results demonstrated differences in recognition performance between numerical components and positional terms, and these patterns varied between the two systems under the tested bilingual conditions.

A key finding of this study was the influence of bilingual speech patterns on voice recognition performance. Although partial recognition was observed, the relatively low complete-match rates for full phrases indicate that the current system may not yet be sufficient for direct clinical

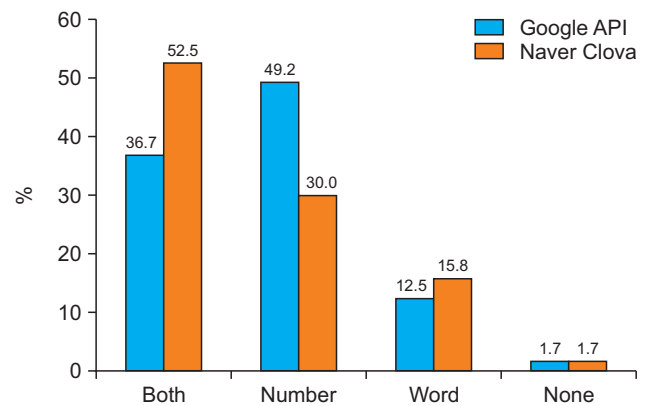


Figure 4. Results of voice recognition for periodontal chart phrases consisting of numerical values and positional terms, shown as percentages for two API models. Outcomes were categorized into four groups according to whether the numerical component or the word component was matched: both matched, only numbers matched, only words matched, and neither matched.

documentation without additional correction mechanisms. Accordingly, these findings should be interpreted as an early signal of feasibility rather than evidence of immediate clinical applicability. Although both APIs have been reported to achieve recognition accuracies exceeding 90% in single-language settings, overall performance in the present evaluation was lower, with an average complete phrase recognition rate of approximately 45% [13]. Recognition errors involving English positional terms frequently reflected misclassification as phonetically similar Korean words. These findings suggest that bilingual pronunciation patterns may affect speech recognition output. Under such conditions, recognition behavior differed from previously reported default performance, for which inconsistent results have been reported regarding the relative performance of the two systems [13].

Clinical environmental factors were also relevant to feasibility. Ambient noise levels in the dental clinic averaged approximately 60 dB and were not substantially influenced by activity at adjacent dental chairs. However, some variability in noise levels was observed, and only scaling procedures were evaluated in this study. Other dental procedures involving rotary instruments may produce greater acoustic variability, underscoring the importance of microphone selection and placement when implementing voice recognition systems in dental settings. Inclusion of all measurements without outlier exclusion did not materially alter the overall noise level, suggesting that the reported values are representative of typical clinical conditions.

Voice recognition has been increasingly investigated as a tool for EMR documentation across medical fields. Previous

studies have reported potential benefits, including reduced documentation burden and clinician burnout [14], as well as improved documentation efficiency in acute care settings [15]. However, many previous investigations have been conducted under controlled or nonclinical conditions, with limited consideration of real-world clinical environments.

Speech-to-text technology presents inherent challenges because it requires the integration of acoustic signal processing and natural language understanding [16]. Although commercial platforms support a wide range of languages, recognition performance may be affected by dialectal variation and bilingual speech input. Recent studies have focused on adapting speech recognition systems to specific linguistic contexts, including dialectal variation in English [17], Korean [18], Chinese [19], and Indian languages [20].

In the Korean medical context, scientific and medical terminology is commonly expressed in English, whereas numerical values are often spoken in Korean [21]. As a result, clinical speech frequently involves mixed-language input rather than a single standardized language. Separating languages within a single spoken phrase presents a substantial challenge for speech recognition systems, particularly when English terms are pronounced with Korean phonetic influence. This issue is not unique to dentistry and may also be relevant in other non-English-speaking healthcare environments.

Previous research on bilingual speech recognition has indicated that training requirements are substantially greater than those for single-language systems [22,23]. Although the development of fully generalized bilingual models for all linguistic combinations may be impractical, feasibility may be improved through localized, task-specific approaches. In dental charting, the vocabulary and structural patterns are relatively constrained. For example, periodontal probing records predictably consist of numerical values and a limited set of positional terms. Leveraging these structured input formats, incorporating correction intervals, or applying constrained vocabularies may improve recognition performance. Recent advances in model-training strategies, including generative modeling and hybrid approaches, have demonstrated improved performance with limited datasets [24,25].

An additional feasibility consideration is the need to distinguish speech intended for chart documentation from routine verbal communication with patients. Implementing trigger words or defined recording intervals may help prevent unintended transcription during clinical interactions.

The current evaluation framework provides a coarse categorization of recognition outcomes and does not fully dis-

tinguish clinically meaningful error types, such as incorrect tooth number, probing depth, or positional term. In addition, numerical recognition was assessed on the basis of complete matching of three-digit sequences, which does not capture digit-level error patterns. Although detailed digit-wise results are partially presented in Supplementary Table S2, a more comprehensive component-level error analysis would be necessary to better assess clinical risk and practical applicability.

This study has several limitations. The evaluated speech inputs were limited to periodontal probing depth phrases and did not encompass the full range of periodontal charting elements. In addition, speech samples were obtained from a limited number of male speakers within a single clinical environment, which may limit the generalizability of the findings. Speaker-dependent variability in speech recognition performance was not formally analyzed, and only one model was evaluated for each microphone type. Further feasibility studies incorporating more diverse speakers, speaker-level analyses, varied clinical environments, and additional speech recognition APIs are warranted.

This study evaluated the feasibility of voice recognition-based EMR documentation for periodontal probing under Korean-English bilingual conditions in a dental clinical environment. Recognition performance differed between numerical and linguistic components, indicating that bilingual speech patterns may influence voice recognition outcomes. These findings suggest that linguistic and environmental characteristics specific to dental practice should be considered when selecting and implementing voice recognition systems for dental EMR documentation.

## Conflict of Interest

No potential conflict of interest relevant to this article was reported.

## Acknowledgments

This research was supported by the 2021 Yonsei University College of Dentistry student research grant.

## ORCID

Young Woo Kim (<https://orcid.org/0000-0003-2646-5140>)

Jin Hyeok Kook (<https://orcid.org/0009-0005-6163-3367>)

Yiseul Choi (<https://orcid.org/0000-0001-6283-9686>)

Wonse Park (<https://orcid.org/0000-0002-2081-1156>)

## Supplementary Materials

Supplementary materials can be found via <https://doi.org/10.4258/hir.2026.32.2.118>.

## References

- Simon L, Obadan-Udoh E, Yansane AI, Gharpure A, Licht S, Calvo J, et al. Improving oral-systemic health-care through the interoperability of electronic medical and dental records: an exploratory study. *Appl Clin Inform* 2019;10(3):367-76. <https://doi.org/10.1055/s-0039-1688832>
- Holroyd-Leduc JM, Lorenzetti D, Straus SE, Sykes L, Quan H. The impact of the electronic medical record on structure, process, and outcomes within primary care: a systematic review of the evidence. *J Am Med Inform Assoc* 2011;18(6):732-7. <https://doi.org/10.1136/amia-jnl-2010-000019>
- Yan Q, Jiang Z, Harbin Z, Tolbert PH, Davies MG. Exploring the relationship between electronic health records and provider burnout: a systematic review. *J Am Med Inform Assoc* 2021;28(5):1009-21. <https://doi.org/10.1093/jamia/ocab009>
- Zick RG, Olsen J. Voice recognition software versus a traditional transcription service for physician charting in the ED. *Am J Emerg Med* 2001;19(4):295-8. <https://doi.org/10.1053/ajem.2001.24487>
- Bergeron B. Voice recognition and medical transcription. *MedGenMed* 2004;6(3):54.
- Cho DB, Lee YR, Lee W, Lee ES, Lee JH. Analyzing health information technology and electronic medical record system-related patient safety incidents using data from the Korea Patient Safety Reporting and Learning System. *Qual Improv Health Care* 2021;27(2):57-72. <https://doi.org/10.14371/QIH.2021.27.2.57>
- Lee W, Seong JJ, Ozlu B, Shim BS, Marakhimov A, Lee S. Biosignal sensors and deep learning-based speech recognition: a review. *Sensors (Basel)* 2021;21(4):1399. <https://doi.org/10.3390/s21041399>
- Hockstein MA, Pope SN, Donnawell K, Chavez SA, Bhat L. Emergency medicine residents on electronic medical records: perspectives and advice. *Cureus* 2019; 11(2):e4027. <https://doi.org/10.7759/cureus.4027>
- Corbit C. When life throws you lemons, make lemonade: voice recognition. In: HIT or miss. 3rd ed. New York (NY): Productivity Press; 2019. pp. 131-4. <https://doi.org/10.4324/9780429031403-8>
- Vine DG, Van Heerden MV, Gerntholtz T. Electronic medical records: why should you consider implementing an EMR? *CME* 2007;25(1):24-8.
- Suvalsky S, Boelman V. Voice recognition technology vs. manual transcription: is voice recognition technology superior to manual transcription for surgical pathology--in terms of return on investment?: a formal study says the answer is no. *MLO Med Lab Obs* 2012;44(9):36, 38-9.
- Strahan RH, Schneider-Kolsky ME. Voice recognition versus transcriptionist: error rates and productivity in MRI reporting. *J Med Imaging Radiat Oncol* 2010;54(5): 411-4. <https://doi.org/10.1111/j.1754-9485.2010.02193.x>
- Choi SJ, Kim JB. Comparison analysis of speech recognition open APIs' accuracy. *Asia Pac J Multimed Serv Converg Art Humanit Sociol* 2017;7(8):411-8. <https://doi.org/10.35873/ajmahs.2017.7.8.038>
- Myers JM. The role the electronic medical record plays in the crisis of provider burnout, its implications and possible solutions. *Univ Lynchburg DMSc Dr Project Assign Repos* 2021;3(3):82.
- George SS. Emergency medical service EMR-driven concept extraction from narrative text [thesis]. West Lafayette (IN): Purdue University; 2021.
- Chandolikor N, Joshi C, Roy P, Gawas A, Vishwakarma M. Voice recognition: a comprehensive survey. *Proceedings of 2022 International Mobile and Embedded Technology Conference (MECON); 2022 Mar 10-11; Noida, India.* p. 45-51. <https://doi.org/10.1109/MECON53876.2022.9751903>
- Kaur A, Sachdeva R, Singh A. Automatic speech recognition in English language: a review. In: Kadyan V, Singh A, Mittal M, Abualigah L, editors. *Deep learning approaches for spoken and natural language processing.* Cham, Switzerland: Springer; 2022. p. 127-45. [https://doi.org/10.1007/978-3-030-79778-2\\_8](https://doi.org/10.1007/978-3-030-79778-2_8)
- Yoo HJ, Seo S, Im SW, Gim GY. The performance evaluation of continuous speech recognition based on Korean phonological rules of cloud-based speech recognition open API. *Int J Netw Distrib Comput* 2021;9:10-8. <https://doi.org/10.2991/ijndc.k.201218.005>
- Peng L, Zhu X, Zhang P. Machine learning-based speech recognition of Chinese dialects method for mobile forensics. *Proceedings of 2021 13th International Conference on Communication Software and Networks (ICC-SN); 2021 Jun 4-7; Chongqing, China.* p. 332-6. <https://doi.org/10.1109/ICCSN52437.2021.9463649>
- Bhardwaj V, Kukreja V, Kaur N, Modi N. Building an

- ASR system for Indian (Punjabi) language and its evaluation for Malwa and Majha dialect: preliminary results. Proceedings of 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT); 2021 Jul 6-8; Kharagpur, India. p. 1-5. <https://doi.org/10.1109/ICCCNT51525.2021.9579471>
21. Son JW, Chang H. Multiculturalism in current and future mathematics teacher education in South Korea. In: Khine MS, Liu Y, editors. Handbook of research on teacher education. Singapore: Springer; 2022. p. 551-71. [https://doi.org/10.1007/978-981-16-9785-2\\_27](https://doi.org/10.1007/978-981-16-9785-2_27)
  22. Neil DE, Andreason T. Examination of voice recognition system to function in a bilingual mode (Report No. NPS5581004). Monterey (CA): Naval Postgraduate School; 1981.
  23. Gonzalez GM, Costello CR, La Tourette TR, Joyce LK, Valenzuela M. Bilingual telephone-assisted computerized speech-recognition assessment: is a voice-activated computer program a culturally and linguistically appropriate tool for screening depression in English and Spanish? *Cult Divers Ment Health* 1997;3(2):93-111. <https://doi.org/10.1037/1099-9809.3.2.93>
  24. Wang J, Kim S, Lee Y. Speech augmentation using WaveNet in speech recognition. Proceedings of 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2019 May 12-17; Brighton, UK. p. 6770-4. <https://doi.org/10.1109/ICASSP.2019.8683388>
  25. Hassine M, Boussaid L, Massaoud H. Tunisian dialect recognition based on hybrid techniques. *Int Arab J Inf Technol* 2018;15(1):58-65.