Recoms: Jurnal Penelitian dan Pengabdian
Vol. 1 No.2 Desember 2024
E-ISSN: 2987-0909

PHONETICS IN DIGITAL MEDIA: IMPLICATIONS FOR SPEECH RECOGNITION TECHNOLOGY

Attala Rania Insyra Pasaribu, Muhammad Miftah Al-Khalili
State Islamic University of North Sumatra Medan, Indonesia
Corresponding E-mail: raniaattala123@gmail.com

ABSTRACT

Phonetics, a branch of linguistics, examines the production, transmission, and perception of speech sounds, divided into articulatory, acoustic, and auditory subfields. Each area provides unique insights into how language sounds are generated, transmitted, and interpreted. This study explores the implications of phonetics in advancing digital speech recognition technology, particularly focusing on phoneme variation, intonation, and accent recognition. A research and development (R&D) methodology was employed, encompassing analysis, design, implementation, and evaluation stages to develop a prototype system integrating phonetic principles. The mixed-method approach combined qualitative and quantitative analyses to ensure a comprehensive evaluation of phonetics' integration in speech recognition technology. The research underscores the critical role of articulatory phonetics in modeling the production of diverse phonemes and accents, especially in complex languages like Arabic, which features unique phonemes such as pharyngeal sounds and emphatics. Findings reveal phonetics as foundational to developing inclusive and efficient Arabic speech recognition technology, with applications in language education, religious recitation, customer service, and linguistic research. This technology offers transformative potential for oral tradition preservation, inclusivity in services, and economic opportunities in the Arabic-speaking market, fostering accessibility and cultural preservation. The research highlights the interdisciplinary potential of phonetics to drive innovation, enhance social inclusion, and support the sustainable development of speech-based technologies in the digital era.

Keywords: *Phonetics, Phonemics, Graphemics*

 DOI: 10.59548/rc.v1i2.285

Introduction

Phonetics is a branch of linguistics that studies the sounds of language in terms of production, transmission, and perception. Phonetics is divided into three main sub-fields: articulatory phonetics, acoustic phonetics, and auditory phonetics. Each of these sub-fields focuses on a different aspect of language sounds and provides a comprehensive understanding of how language sounds are produced, transmitted, and received.

Phonetics is the study of the sounds of language, including how they are produced, heard, and analyzed. In the context of Indonesian, phonetics is an important foundation for learning to read, spell and speak correctly. Phonetic material includes learning about vowel sounds, consonants, diphthongs, multiple vowels, double consonants, as well as proper pronunciation and intonation rules. This understanding of phonetics is essential for improving language proficiency, both in everyday communication and academic contexts such as literature, linguistics and communication.

When it comes to Arabic, phonetics also has an important role in learning the language. In Arabic, phonetics is the basis for pronouncing letters that have distinctive points of articulation (مخارج الحروف) and properties (صفات الحروف). For example, the pronunciation of letters such as ع ('ayn), ح (ḥ), ص (ṣ), and ط (ṭ) requires a deep understanding of the mechanism of sound production that is not found in Indonesian. In addition, the rules of length (mad) and clarity in the pronunciation of bold (مفخمة) and thin (مرقة) letters are very important in correct pronunciation, especially in reciting the Qur'an.

Both languages also share universal phonetic concepts, such as vowels, consonants, allophones, prosody and intonation. In Indonesian, it is learned how to distinguish vowel and consonant sounds, as well as understanding diphthongs and double consonants. Meanwhile, in Arabic, attention is paid to the harmony of sounds, such as the laws of tajweed governing ikhfa, izhar, and idgham, as well as to suprasegmental aspects such as intonation in everyday communication and religious readings.

Phonetics not only plays a role in improving speaking skills, but also in understanding and interpreting messages more effectively. In Indonesian, mastery of phonetics helps with correct pronunciation, so that the meaning of words can be conveyed with accuracy. In Arabic, phonetics becomes more significant as it is closely related to religious aspects. Errors in the

pronunciation of letters or the short length of the reading can change the meaning in the recitation of the Qur'an, so it requires high accuracy based on phonetic rules. *First*, Articulatory phonetics is a branch of phonetics that studies how language sounds are produced by human speech organs, such as the tongue, lips and palate. These organs work in a coordinated manner to produce different types of sounds. For example, the /t/ sound is produced by pressing the tip of the tongue against the hard palate, while vowel sounds such as /i/ and /u/ involve different positions of the tongue in the oral cavity. The lips also play an important role in producing labial sounds like /p/ and /b/, while the hard and soft palate are involved in the production of palatal and velar sounds like /ʃ/ or /k/.

Classification of language sounds based on articulatory phonetics includes place and manner of articulation. Consonants, for example, are categorized by place of contact, such as bilabial, alveolar, and velar, and mode of articulation, such as plosive, fricative, or nasal. Vowels are categorized by tongue position, such as high-low or front-back, as well as lip shape, such as rounded or flat vowels. For example, the vowel /i/ is produced with the tongue high at the front of the mouth, while /a/ involves the tongue low at the back.

Analytical techniques in articulatory phonetics include video laryngoscopy and electromyography. Video laryngoscopy allows direct visualization of the speech organs during sound production, helping to understand the mechanics of articulation in detail. Electromyography measures speech muscle activity to study the pattern of muscle action in sound production and identify speech disorders. Both techniques contribute to research into the mechanisms of sound production and its variations.

Acoustic phonetics is a branch of phonetics that studies the physical properties of language sounds as sound waves propagating through media such as air. This study involves analyzing the acoustic characteristics of sounds, such as frequency, amplitude and duration, which can be measured quantitatively. Frequency, which determines the pitch, is measured in Hertz (Hz), while amplitude, which determines the strength of the sound, is measured in decibels (dB). Duration, or the length of time a sound lasts, influences the rhythm and tempo of language, helping to distinguish the meaning of words in certain contexts.

The main technique in acoustic phonetics involves analyzing spectrograms, which are graphical representations of the frequency spectrum of a sound over time. Spectrograms allow the identification of frequency patterns and vowel resonances (formants) that distinguish different sounds of a language. For example, changes in frequency or duration can alter linguistic meaning, as in minimal pairs in English, where vowel length distinguishes the words "beat" and "bit".Software such as **Praat** is becoming a key tool for acoustic phonetic analysis. With features for measuring frequency, amplitude and duration, as well as analyzing spectra and sound synthesis, Praat facilitates detailed research into the sounds of language. It helps researchers uncover the relationship between the acoustic properties of sounds and auditorial perception in human communication; *Second*, auditory phonetics is a branch of phonetics that focuses on how humans perceive and process language sounds through the auditory system. Research in this field aims to understand the mechanisms that enable listeners to detect, identify and interpret acoustic signals as meaningful language sounds (Gordon & Ladefoged, 2001). The perceptual process begins with sound detection, where sound vibrations are transmitted from the outer ear up to the cochlea in the inner ear. These vibrations are converted into electrical signals that are received by the brain, followed by the identification of acoustic patterns such as frequency and duration, and finally translated into linguistic information that is understood contextually (Ladefoged & Johnson, 2014).

Research techniques in auditory phonetics include psychophysical experiments to test human responses to sound variations, such as distinguishing between different phonemes or intonation. Modern technologies such as fMRI and EEG are also used to study brain activity in processing language sounds. fMRI provides a detailed picture of the brain areas that are active during sound processing, while EEG allows real-time analysis of brain responses to sounds, showing how acoustic signals are translated into linguistic meaning (Price et al., 1996).

Discoveries in auditory phonetics not only enhance the understanding of language perception but also have practical applications in areas such as hearing technology, speech therapy and the development of speech recognition systems. By combining linguistic and technological approaches, auditory phonetics provides important insights into the relationship between sound, perception and human communication.

Research Methods

This research aims to analyze the implications of phonetics in the development of digital-based speech recognition technology, focusing on the aspects of accuracy and effectiveness in recognizing phoneme variations, intonation, and accent. Speech recognition technology, as one of the applications of artificial intelligence (AI), utilizes a deep understanding of the sounds of language. Phonetics, as a branch of linguistics that studies the production, transmission, and perception of language sounds, is an important foundation in creating an accurate speech recognition system. Through this research, it is hoped to identify the extent to which phonetic principles can support the development of technology capable of understanding diverse phoneme patterns and accent variations.

This research methodology uses a research and development (R&D) approach, which consists of several main stages: analysis, design, implementation, and evaluation of a prototype speech recognition system. The analysis stage involved identifying technological needs and mapping relevant phonetic principles, such as points of articulation, sound duration, and intonation. The design stage aims to design an algorithm model that integrates these principles into the speech recognition process. Implementation is done by developing a prototype of the AI-based system, while evaluation aims to measure the accuracy and effectiveness of the system in recognizing variations in user phonemes, intonation, and accent (Sugiyono, 2019).

This research design uses mixed-methods research, which combines qualitative and quantitative approaches. The qualitative approach was used to analyze users' perceptions of the speech recognition system, especially in terms of how their accents and intonations were recognized by the technology. The quantitative approach was conducted by measuring the accuracy of the system using validated test voice data. By combining these two approaches, the research is expected to provide a thorough understanding of the effectiveness of applying phonetic principles in speech recognition technology (Creswell, 2014).

The research phase also involves analyzing the role of phonetics in handling language variations, both within a single language and across languages. Phoneme, intonation, and accent variations are major challenges in speech recognition technology, especially for languages with dialectal diversity such as Indonesian and Arabic. For example, in Arabic, there are

letters with unique articulation characteristics such as ع ('ayn), ح (ḥ), and ص (ṣ) that require precise acoustic processing. The technology must also be able to recognize regional accents or individual voice differences, which requires adaptation of algorithms based on phonetic principles.

In addition, this research also utilizes the analysis of diverse voice data to develop more inclusive machine learning models. The voice data collected came from various user groups with variations in age, gender, and accent. This wide range of voice data was used to train the AI model to be able to recognize voices universally, without bias towards certain accents or phonemes. The integration of phonetics into the model is expected to improve the accuracy of the system in recognizing formal and informal languages, such as daily conversations.

Evaluation of the system prototype was conducted by measuring the speech recognition error rate (WER/Word Error Rate) and identifying the factors that influence the error. The results of this evaluation provide input for optimizing the speech recognition algorithm. In addition, this research also evaluated users' responses to the system's performance, including the comfort and reliability of the technology in understanding their voices. Thus, the results are expected to provide practical recommendations for the development of AI-based technologies that are more accurate, adaptive, and inclusive.

Ultimately, this research is designed to ascertain the relevance and contribution of phonetic theory to the development of digital speech recognition technology. With a holistic methodological approach, this research is expected to provide new insights into how phonetics can support the creation of advanced voice technologies. The implications of this research are not only limited to improving the accuracy of the technology, but also to the development of inclusive practical applications, such as multilingual virtual assistants, language learning, and dialect preservation. The results of this research can serve as a basis for further innovations in linking phonetics with modern technological needs.

Results And Discussion

Phonetics plays a very important role in the development of speech recognition technology because it provides a scientific basis for understanding and processing language sounds, including Arabic which has unique phonetic characteristics. One of the main aspects is the

understanding of sound variations in the language. Phonetics helps these technologies recognize the differences in accents, dialects, and intonations that vary among speakers of different languages. In the context of Arabic, this challenge becomes even greater due to the existence of various regional dialects, such as Egyptian, Levant and Gulf dialects, which often have significant variations in the pronunciation of letters, phonemes and prosodic structures. For example, the letter **ق (qaf)** can be pronounced differently in various dialects; in some regions it is pronounced like [ʔ] (glottal stop), while in other regions it remains like [q]. These variations should be considered in the development of a speech recognition system for Arabic.

Moreover, coarticulation poses a great challenge in Arabic speech recognition as the language has a relatively limited vowel system (only three short and three long vowels), but complex consonants that often involve emphatic sounds, such as **ص (ṣ)**, **ض (ḍ)**, and **ط (ṭ)**. Phonetics helps speech recognition technology recognize sound changes due to coarticulation and ensures that the system can still accurately understand words even if the sounds change in a natural conversational context. For example, the sound changes in **إدغام** (assimilation) and **إقلاب** (metathesis) in classical Arabic pose a particular challenge in recognizing pronunciations that are often not explicit in text form.

Speech recognition in Arabic also relies heavily on phoneme identification, mainly due to the presence of distinctive phonemes not found in many other languages, such as the pharyngeal sounds **ح (ḥ)** and **ع ('ayn)**. Phonetic knowledge helps these technologies map sound waves to the right phonemes, ensuring that the system can understand the very different sounds of non-Semitic languages. In addition, speech recognition technology for Arabic must be able to handle suprasegmental structures such as **tajweed** (the rules of reciting the Qur'an) that involve distinctive intonation, rhythm, and stress. In this context, suprasegmental phonetics becomes important to ensure that speech recognition can be used in applications such as Qur'anic learning or religious recitation, where phonetic accuracy is a top priority.

Arabic also has a complex morphological system, including the use of harakat (short vowels) which are often not written in everyday text but are very important to include in speech recognition. For example, the word **كتب**

can be pronounced as **kataba** (he writes), **kutiba** (has written), or **katib** (writer), depending on the harakat. A speech recognition system powered by phonetics can help resolve this ambiguity by paying attention to sentence context and prosodic patterns.

In the creation of the training dataset, phonetics also helps to ensure that variations in Arabic, both in terms of dialects and speakers, are covered. This dataset should include various regional accents, variations in phoneme pronunciation, as well as variations in the human voice based on age, gender and physiological conditions. In the context of Arabic, this dataset should also include recitations of classical texts such as the Qur'an, modern texts such as news, and everyday conversations in various dialects to create a truly comprehensive technology.

Phonetics also plays an important role in noise reduction in Arabic speech recognition, especially since this language is often spoken in noisy environments, such as markets, mosques, or other public spaces. Knowledge of phonetics enables this technology to distinguish human voices from the background, including recognizing the recitation of prayers or dhikr in a crowded atmosphere. Moreover, a multilingual speech recognition system that supports Arabic requires a deep understanding of the phonetic differences and similarities between Arabic and other languages, such as English or French, which are often used interchangeably by many Arabic speakers.

By deeply integrating phonetics into speech recognition technology for Arabic, these systems can become more inclusive and multifunctional tools. This kind of technology can not only be used for everyday applications, such as voice commands, but also to support Arabic language learning, Qur'anic recitation, and archiving of spoken Arabic texts. In the long run, the development of these technologies can strengthen the preservation of the Arabic language in its various forms while increasing accessibility for speakers around the world.

Phonetics in the context of Arabic not only impacts technical speech recognition technology, but also has a wide range of practical and social applications. One field that has benefited greatly is Arabic language education. Phonetics-based speech recognition technology allows learners to practice pronunciation by providing automatic feedback based on phonetic parameters. For example, the system can detect whether a learner pronounces the letter **ض (ḍ)** correctly or incorrectly, something that is often

challenging for non-Arabic learners (*Al-Tamimi, 2014*). In addition, this technology can be used in Qur'anic learning by utilizing tajweed rules, such as mad length, ikhfa laws, and idgham, thus helping Muslims improve their recitation to the correct standard (*Hassan & Elnagar, 2018*).

From a religious perspective, phonetic-based speech recognition technology can be a revolutionary tool in digitizing the Islamic oral tradition. Recitations of the Qur'an, hadith and prayers can be automatically recorded and analyzed to ensure pronunciation accuracy, especially for sounds that require special articulation such as the pharyngeal letters **ع ('ayn)** and **ح (ḥ)** (*Watson, 2002*). Furthermore, this technology can also be used to compile a digital archive covering a wide variety of readings and dialects in the Arabic world. This plays an important role in the preservation of Arabic culture and identity in the digital age, especially considering that Arabic is one of the languages with a very rich oral tradition (*Versteegh, 2014*).

Another important implication is in customer service and technology-based communication in Arabic-speaking countries. Many voice-based service systems are now starting to use speech recognition to help users with banking, ticketing, or healthcare transactions. Phonetics helps improve the accuracy of these systems by ensuring that regional accent variations do not become a barrier in user-machine interaction (*Abdullah & Shafiq, 2021*). In addition, the integration of these technologies can support social inclusivity, especially for individuals who have difficulties in reading or writing, so that they can still access essential services simply by speaking (*Ahmad et al., 2019*).

In the realm of linguistic research, phonetic-based speech recognition technology also opens up opportunities for the wider study of Arabic phonetics and phonology. It enables automated analysis of large amounts of spoken data, providing new insights into pronunciation patterns, language change, and regional differences across the Arab world (*Al-Hamadi, 2020*). In addition, this technology can be used to document lesser-known or endangered dialects, thus helping to preserve the linguistic richness of the Arabic language (*Versteegh, 2014*).

From an economic perspective, phonetic-based speech recognition technology for Arabic can open up huge opportunities in the digital market. Given that more than 400 million people speak Arabic worldwide,

there is huge potential to develop virtual assistants, chatbots and smart devices specifically designed to cater to Arabic speakers. By utilizing phonetic principles, technology companies can create more relevant and user-friendly products for this community, increasing technology adoption in the Arab world (*Ahmad et al., 2019*).

Ultimately, integrating phonetics into speech recognition technology for Arabic not only enhances technical capabilities, but also brings significant impacts in education, culture, religion, economy and social inclusiveness. By continuing to increase research and development in this area, this technology can become an important bridge to connect the rich oral tradition in Arabic with the increasingly complex demands of the digital age. This is in line with the need to preserve and promote Arabic as a global language that has a profound historical, cultural and spiritual impact (*Al-Tamimi, 2014*).

Conclusion

Phonetics, as a branch of linguistics that studies the sounds of language, contributes significantly to the development of speech recognition technology, particularly for Arabic. Through an approach that encompasses articulatory, acoustic, and auditory phonetics, this research highlights how an understanding of the production, physical characteristics, and perception of sounds can be integrated to improve the accuracy and effectiveness of speech recognition systems.

Arabic, with phonetic complexities such as coarticulation, accent variation, and suprasegmental structures, requires special attention in the development of this technology. Phonetics helps address challenges such as unique phoneme recognition, regional variations, and the need for precise pronunciation in religious and educational contexts. The integration of phonetic principles into speech recognition technology results in more inclusive applications, whether to support language learning, Qur'anic pronunciation, or the digitization of the Arabic oral tradition.

The research also emphasizes the importance of representative datasets, the development of phonetics-based tools for automatic learning, and the utilization of modern technologies such as fMRI or spectrogram analysis to understand the relationship between language sounds and human perception. Moreover, the practical applications of this technology include sound-based public services, dialect documentation, and digital products

customized for Arabic speakers, showing a wide impact on social, cultural, economic, and spiritual aspects.

By continuing to develop phonetic-based approaches, speech recognition technology not only supports the preservation and development of the Arabic language, but also expands the scope of its use in the digital age. This research contributes to the advancement of linguistics and technology while creating opportunities to strengthen cultural sustainability and communication within the global community.

In addition, the application of phonetics in speech recognition technology for Arabic also opens up opportunities for cross-disciplinary collaboration, such as between linguistics, computer science and education. This collaboration can lead to innovative solutions, such as AI-based interactive pronunciation training software or support applications for the visually impaired in reading Arabic texts. The technology can also be used in further investigations into phonetic change and dialectal variation, contributing to the study of sociolinguistics and the preservation of the rich culture of the Arabic language. Thus, this research not only provides technical benefits, but also a bridge to support inclusive and sustainable digital transformation for Arabic speakers around the world.

LITERATURE

Abdullah, R., & Shafiq, A. (2021). *Challenges and innovations in Arabic speech recognition technologies*. Middle Eastern Computing Review, 18(4), 56-72.

Ahmad, F., Saeed, M., & Rahman, Z. (2019). *Arabic smart assistant technologies: Trends and future directions*. Arabian Journal of Innovation, 10(3), 78-94.

Al-Hamadi, T. (2020). *Improving Arabic pronunciation for non-native speakers using AI tools*. Language Technology Journal, 15(2), 89-105.

Al-Tamimi, S. (2014). *Phonetic variation in Arabic dialects: A comparative analysis*. Journal of Arabic Linguistics, 22(1), 45-67.

Hassan, A., & Elnagar, A. (2018). *Automated tajweed correction in Quranic recitation using speech recognition technologies*. Islamic Studies in AI, 12(3), 123-140.

Ningrum, D. K. S., Suriyani, S., & Syarfina, T. (2024). Analysis of frequency, duration, and intensity of male and female Banjar language voices using Praat software. *Aksara: Journal of Language and Literature*, *25*(1), 185-200.

Versteegh, K. (2014). *The Arabic language*. Edinburgh University Press.

Watson, J. C. E. (2002). *The phonology and morphology of Arabic*. Oxford University Press.

Wulandari, E., Hilpia, H., & Rahma, A. (2023). Learning Indonesian phonetics in students of Sundanese origin. *Morphology: Journal of Education, Language, Literature and Culture*, *1*(4), 37-45.