DOI: 10.61782/fa.2025.0302

FORUM ACUSTICUM EURONOISE 2025

DIGITAL VOICES: AN ANALYSIS OF SPEECH INTELLIGIBILITY THROUGH MOBILE DEVICES

Pérez Vargas, Rocío[1*] **Rodríguez Montaño, Víctor M**[1]
[1] University Institute of Research in Applied Linguistics, University of Cádiz, Cádiz, Spain

ABSTRACT

The phonetic-acoustic interface is crucial in the forensic domain, yet its study and practical application have been relatively underexplored. Its significance stems from the ability to analyze vocal characteristics within legal contexts, providing valuable insights for tasks such as speaker identification, voice authentication and the assessment of message quality. This interface also studies how voice features are transmitted across various settings, contributing to a deeper understanding of the clarity of communicative exchanges.

In this regard, this study seeks to emphasize the critical role of mobile devices in determining speech intelligibility. Specifically, it aims to analyze how different devices, including prominent brands such as iPhone, Xiaomi, Samsung and Motorola, affect the clarity of oral messages. The underlying motivation of this research is to identify which brands enhance speech intelligibility and which hinder it, thereby influencing the effectiveness of communication in mobile phone interactions. To achieve this, ArtemiS SUITE software will be used, a specialized tool for the objective analysis of speech intelligibility. The findings from this analysis will provide valuable insights into the relationship between mobile technology and speech perception, contributing to the development of more effective communication tools.

Keywords: *forensic acoustics, mobile devices, speech intelligibility, vocal characteristics.*

[*]***Corresponding author***: *rocio.perez@uca.es*

1. INTRODUCTION

In recent decades, the emergence of information and communication technologies has profoundly reshaped the ways in which human interactions are carried out. Within this evolving landscape, mobile telephony has become the dominant mode of communication, surpassing traditional channels and facilitating the instantaneous exchange of information. Although initially designed for voice communication, mobile platforms such as WhatsApp and Telegram have gained significant popularity, with voice messaging increasingly replacing conventional phone calls.

This technological advancement has not only facilitated greater global connectivity but has also introduced challenges concerning the quality and reliability of transmitted spoken messages. Such reliability is particularly critical in fields where the accuracy and integrity of information are of utmost importance, such as forensic analysis. The widespread use of mobile phones raises additional concerns regarding the fidelity of verbal communication. Variations in hardware configurations and audio processing algorithms among different manufacturers can lead to significant differences in how spoken messages are perceived. These discrepancies are particularly evident in the quality of microphones and speakers, and the way each device processes and compresses audio signals. Therefore, it is crucial to examine how these factors impact speech clarity and intelligibility, especially in contexts where the precision of information is vital, such as in forensic applications.

The phonetic-acoustic interface, defined as the study and interpretation of vocal characteristics and the context in which oral communication occurs, plays a crucial role in this field. In this context, this interface facilitates the analysis of voice parameters, including fundamental frequency and intensity, which are essential for research in areas such as speaker identification, recording authentication, and the assessment of message quality in legal contexts.

The present study, situated within this interface, aims to elucidate the key role of mobile devices in influencing speech intelligibility. Specifically, the research seeks to examine the impact of the inherent characteristics of various devices, focusing particularly on widely recognized brands such as iPhone, Samsung, and Xiaomi, on the clarity and accuracy of transmitted messages. This investigation is motivated by the need to understand the significant influence of technological differences on speech transmission and, consequently, on the quality of acoustic evidence in legal proceedings.
To address this issue, a study has been designed that integrates a multidisciplinary approach, combining elements of acoustic engineering and forensic analysis. The rationale behind this study lies in the recognition that, despite technological advancements, there remains a gap in understanding the interaction between mobile technology and speech perception. In legal proceedings, where the authenticity and accuracy of information are critical, it is essential to comprehend these interactions in order to enhance forensic analysis methods and optimize communication systems. Identifying devices that offer superior clarity in speech transmission will improve the reliability of acoustic evidence and guide the development of future technologies aimed at enhancing intelligibility in mobile communication.
Previous research has demonstrated that, due to the inherent characteristics of mobile phones, both the speaker and hearer of a communication may be situated in different environmental contexts, potentially leading to noise interference and the lack of mutual feedback. This impedes the speaker's ability to adapt their speech style to the communicative context, as would occur in a face-to-face conversation between multiple individuals [1].
Conversely, the evaluation of speech transmission over mobile devices has traditionally focused on speech quality, given the typically high signal-to-noise ratios (SNR) in mobile communications, which generally ensure that speech intelligibility is maintained[1] [2]. However, in noisy environments, such as public transportation, the speech signal can be significantly affected by background noise, even when the hearer is in a quiet setting [2].
The position of the mobile device relative to the user, whether the loudspeaker is active, the use of headphones, or the device being held near the ear, also significantly influences factors such as speech intelligibility and the sound pressure levels (SPL) that reach the ear. These factors are shaped by several variables, including the type of mobile device, its emission power, the morphology of the user's head, the positioning and angle of the device relative to the head, and the shape and posture of the hand holding the phone [3]. In the absence of background noise, Thielens (2019) asserts that speech intelligibility and SPLs are higher when the phone is held close to the ear compared to the use of headphones or a loudspeaker, particularly when background noise is present in the environment [3].
This highlights the importance of mobile devices in ensuring high intelligibility across various communication scenarios, a critical factor in maintaining message fidelity and minimizing misunderstandings in high-stakes contexts, such as forensic analysis or emergency situations.

2. METHODOLOGY

In the present study, an analytical investigation was conducted on four commercially available cell phone brands: iPhone, Xiaomi, Samsung, and Motorola. Specifically, the following models were analyzed:

- iPhone 7 (i7)
- iPhone 12 (i12)
- iPhone 15 Pro (i15)
- Samsung Galaxy A33 (SG)
- Motorola Edge 40 Neo (ME)
- Xiaomi Redmi Note 8 Pro (X8)
- Xiaomi Redmi Note 12 Pro (X12)

The GRAS 45BC KEMAR Head and Torso Simulator (GRAS Sound & Vibration, Holte, Denmark) was used to ensure the reliability of the recordings, considering the anatomical structure of the human head and torso. The simulator was positioned in an acoustically controlled environment, free from background noise, and recordings were made using only the microphone located on the right ear of the simulator.
MOTU 8PRE and MOTU 828 ES audio interfaces (MOTU, Cambridge, Massachusetts, USA) were employed to ensure proper capture and processing of the audio signals, as well as to facilitate interconnection between the various acoustic equipment, ensuring faithful and lossless signal transmission.
Given the variety of potential interactions between the speaker and hearer—for example, both participants using speakerphone mode, the speaker using the speakerphone while the hearer is holding the phone to their ear ("held mode"), the reverse situation, that is, where the speaker holds the device to their ear and the hearer uses the

[1] S. Jørgensen et al. (2015) suggests that, in situations where intelligibility is very low, individuals tend to either end the conversation or change their location to improve the SNR [2].

speakerphone, and finally, the configuration where both participants hold the device to their ear—, the analysis has been simplified by keeping the hearer's mobile device in a fixed position across all cases, specifically using held mode, thus eliminating potential background noise interference. In the case of the hearer, the device was initially placed in held mode at the simulator's right ear (figure 1), with the speaker positioned 15 cm from the ear and the volume set to maximum. In the second configuration, the mobile phone was used in speakerphone mode (figure 2), with the speaker placed in front of the simulator's mouth at a distance of 20 cm from the ear, also with the volume set to maximum.

Figure 1. Hearer's phone position in held mode.

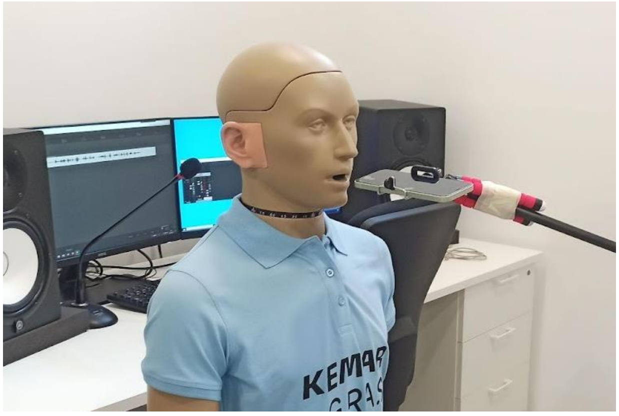

Figure 2. Hearer's phone position in speakerphone mode.

The audio recordings were obtained using REAPER software (Cockos Incorporated, Rosendale, New York, USA), ensuring that each signal was calibrated within a range of -6 to -12 dB to optimize the SNR.

Given the pervasive use of protective drop-resistant cases among the general population, the study was conducted with the devices equipped with these cases.

The evaluation of speech intelligibility was conducted through the implementation of telephone calls between the designated devices under evaluation. The study utilized the iPhone 12 and the Xiaomi Redmi Note 8 Pro as the reference devices for the calls. These two devices were utilized as emission and/or reception points in the speech intelligibility tests, thereby establishing a baseline for the subsequent comparison of the performance of the other models that were analyzed.

The experimental design enabled the determination of the influence of both the speaker and the microphone of each device on message transmission.

The transmitted message consisted of the same sentence spoken by two speakers, a woman and a man, ensuring the identity of the verbal content in all tests. These voices were initially recorded with a KM 150 condenser microphone (Neumann, Berlin, Germany) to establish a speech intelligibility reference.

Speech intelligibility analysis was performed using ArtemiS SUITE software (HEAD acoustics, Herzogenrath, Germany), allowing for a quantitative and objective assessment of the quality of speech transmission on the different devices studied

3. RESULTS

Speech intelligibility, measured from the condenser microphone, considered the reference value due to its fidelity to reality, reached 90% for the male voice and 85% for the female voice.

However, the findings of this study demonstrate that speech intelligibility exhibits substantial variability contingent on the phone usage mode and the device model. The findings underscore the necessity of contemplating two distinct scenarios: firstly, the intelligibility of the devices under scrutiny when operating as hearers, with the reference devices (i12 and X8) functioning as speakers; and secondly, the converse scenario, wherein the reference devices act as hearers and the other devices assume the role of speakers.

In held mode, the intelligibility values for the female voice ranged from 49% to 82% (see Table 1). Maintaining the reference devices as speakers, the X12 model demonstrated the highest values in this condition, reaching up to 74%, while the ME model exhibited the lowest values. In scenarios where the reference devices functioned as hearers, thereby assessing the microphone performance of diverse hearers, the X12 model yielded superior intelligibility.

For the male voice, the results were analogous, with values ranging from 39% to 79% (see Table 2). As was the case in the previous instance, the X12 model demonstrated superior performance in comparison to the other models.

Table 1. Speech Intelligibility of the female voice across different mobile phones in held mode.

		Hearer's mobile						
		i12	X8	i7	i15	SG	ME	X12
Speaker's mobile	i12	-	59	59	65	71	65	70
	X8	70	-	56	64	51	55	74
	i7	56	67	-	-	-	-	-
	i15	49	69	-	-	-	-	-
	SG	59	57	-	-	-	-	-
	ME	62	51	-	-	-	-	-
	X12	72	82	-	-	-	-	-

Table 2. Speech Intelligibility of the male voice across different mobile phones in held mode.

		Hearer's mobile						
		i12	X8	i7	i15	SG	ME	X12
Speaker's mobile	i12	-	50	42	54	51	74	66
	X8	65	-	65	72	67	49	79
	i7	67	60	-	-	-	-	-
	i15	49	75	-	-	-	-	-
	SG	39	39	-	-	-	-	-
	ME	73	55	-	-	-	-	-
	X12	57	65	-	-	-	-	-

Conversely, when the phone was utilized in speakerphone mode, a substantial enhancement in speech intelligibility was observed. This observation was consistent across both female and male voices, as evidenced by the significant increase in values observed in Tables 3 and 4, respectively. In this context, the i7 and X12 models demonstrated the highest values when the reference devices functioned as speakers. Conversely, when the reference devices functioned as hearers, thereby assessing the audio quality emitted by the remaining models, the ME and i16 models exhibited superior speech intelligibility outcomes.

Table 3. Speech Intelligibility of the female voice across different mobile phones in speakerphone mode.

		Hearer's mobile						
		i12	X8	i7	i15	SG	ME	X12
Speaker's mobile	i12	-	75	90	80	64	76	79
	X8	73	-	91	83	71	75	85
	i7	64	70	-	-	-	-	-
	i15	75	81	-	-	-	-	-
	SG	69	74	-	-	-	-	-
	ME	77	82	-	-	-	-	-
	X12	66	73	-	-	-	-	-

Table 4. Speech Intelligibility of the male voice across different mobile phones in speakerphone mode.

		Hearer's mobile						
		i12	X8	i7	i15	SG	ME	X12
Speaker's mobile	i12	-	72	91	80	68	73	82
	X8	69	-	93	79	70	78	87
	i7	60	65	-	-	-	-	-
	i15	77	82	-	-	-	-	-
	SG	67	73	-	-	-	-	-
	ME	80	84	-	-	-	-	-
	X12	68	75	-	-	-	-	-

4. DISCUSSION

In contrast to the results obtained by Thielens et al. (2019) [3], our data suggest greater speech intelligibility when the mobile phone is in speakerphone mode compared to held mode. This finding is particularly relevant since the aforementioned study [3] indicates that, in most cases, speech intelligibility was higher when the device was held close to the ear due to the proximity of the microphone to the speaker. However, it is crucial to note that the methodologies used to obtain these results are inherently different, which may account for the discrepancies observed.

In the study conducted by Thielens et al. (2019) [3], speech intelligibility was assessed using a perceptual test in which participants listened to speech recordings and then transcribed what they had understood. This type of test is commonly used in acoustic research to measure the clarity and comprehensibility of speech under various acoustic and signal conditions, providing a direct evaluation of the listeners' ability to understand the spoken message. The transcription of the recordings allows for the assessment of

how signal characteristics, such as background noise, sound quality, or volume variations, affect the perception of intelligibility.
Conversely, the present study utilizes the Aremis SUITE software, employing an automated and quantitative approach to assess speech intelligibility in accordance with the ASA/ANSI S3.5-1997 (R2024) standard [4]. The utilization of this software facilitates the acquisition of a precise and objective measure of intelligibility, thereby circumventing the potential influence of subjective biases that may be present in perceptual tests conducted by human evaluators.
The main difference between the two approaches lies in the fact that, while the method used by Thielens et al. (2019) [3] relies on human perception, which is subject to variability among individuals, the system employed in our research offers a more standardized and replicable assessment. Although both approaches aim to measure speech intelligibility, the methodological differences may explain the variations in the results obtained, particularly regarding the comparison between different mobile phone usage modes, such as speakerphone mode and held mode.
The impact of the usage mode on speech intelligibility was evident in the results. In the held mode condition, no device reached the expected reference intelligibility values for optimal communication conditions (85% for female voice and 90% for male voice). In contrast, in the speakerphone condition, some devices exceeded these reference values, with recordings of up to 94% for the male voice and 91% for the female voice in the i7 and X12 models. This suggests that speakerphone mode provides a significant advantage in terms of speech signal clarity.
The differences between models and brands were also evident in the results. While greater variations between devices were observed in held mode, intelligibility improved across all devices in speakerphone mode. These findings suggest that certain devices handle speech capture and reproduction in speakerphone mode better than in conventional usage.
Regarding the difference between male and female voices, it was found that, in general, the male voice achieved higher intelligibility values across most devices. However, this difference was reduced in speakerphone mode, where both voices approached the reference values more closely. This could indicate that speakerphone mode favors the transmission of frequencies associated with the female voice, which are typically more affected by audio compression in mobile phones.
The analysis of audio processing in the evaluated devices suggests that none of the phones in held mode achieved the intelligibility levels obtained with the condenser microphone, indicating that the compression and audio processing systems reduce the quality of the transmitted signal. In contrast, some models in speakerphone mode achieved values close to those obtained with reference microphones, suggesting that the audio processing in this condition is less aggressive or that the environment favors speech perception.
While these results reflect measurements in controlled conditions without background noise, it is important to consider that, in real-world situations, speech intelligibility will be lower due to the presence of environmental noise, such as traffic, background conversations, wind, or reverberation in closed rooms.
The speakerphone mode, although beneficial in terms of intelligibility in a controlled environment, may be more susceptible to environmental noise, as the phone's microphone captures a greater amount of external sounds. In situations with echo or reverberation, such as in large rooms or open spaces, the sound played through the speakerphone could degrade more than that perceived in held mode. Furthermore, the presence of noise-canceling algorithms in some phones could affect intelligibility in a variable manner, depending on the type of noise in the environment. In particular, constant noises, such as air conditioning or distant traffic, may be filtered more effectively than intermittent and unpredictable noises, such as honking horns, nearby voices, or strong wind, which can significantly degrade speech intelligibility.

5. CONCLUSION

Despite the limited scope of the present sample, the findings unmistakably indicate that both the utilization mode and the device model exert a substantial impact on speech intelligibility in mobile phones. The methodology employed reveals that the utilization of speakerphone mode enhances speech intelligibility. This phenomenon can be attributed to the fact that manufacturers deliberately design microphones with the understanding that their effectiveness may be hindered in noisy environments. This consideration is probably the impetus behind the design of devices aimed at maximizing sound clarity in noisy situations, either through amplification or improvements in sound capture.
The most straightforward solution to enhance speech intelligibility would be to simply increase the signal level, that is, the device's output volume. While this approach can be effective to some extent, it is limited by the intrinsic characteristics of the speaker and the unpleasant or even painful experience listeners endure due to excessively loud sounds. Consequently, an alternative strategy entails

modifying the time or frequency characteristics of the speech signal while maintaining a restrained output volume. This approach has the potential to enhance speech intelligibility while circumventing the disadvantages associated with elevated sound levels [1].

6. REFERENCES

[1] E. Jokinen: "Techniques for speech intelligibility enhancement in mobile telephony," Ph.D. Thesis, Aalto University, Helsinki, Finland, 2017.

[2] S. Jørgensen, J. Cubick, and T. Dau: "Speech intelligibility evaluation for mobile phone," *Acta Acustica united with Acustica*, vol. 101, pp. 1016-1025, 2015.

[3] A. Thielens, A. Bockstael, S. Declerck, R. Aminzadeh, S. Aerts, D. Botteldooren, et al.: "Mobile phones: A trade-off between speech intelligibility and exposure to noise levels and to radio-frequency electromagnetic fields," *Environmental Research*, vol. 175, 2019.

[4] ASA/ANSI S3.5-1997 (R2024): Methods for Calculation of the Speech Intelligibility Index.