Short Sound Messages for Enhancing Digital Accessibility and Reducing Digital Divide

Manuele Favero*
University of Padova
Padova, Italy

Chiara Schiavo*
University of Padova
Padova, Italy

Mirco Zavarise
University of Padova
Padova, Italy

Leonardo Badia
University of Padova
Padova, Italy

Sergio Canazza
University of Padova
Padova, Italy

Luciano Murrone
Ogenus s.r.l.
Padova, Italy

Abstract

As digital services increasingly rely on visual and touch-based interaction, users with visual or motor impairments are often excluded from seamless access to information. Despite ongoing progress in assistive technologies, accessibility gaps persist across public and private contexts, limiting independent access to digital content. To address this issue, we investigate the potential of TISCODE, a sound-based system that employs Short Sound Messages (SSMs) to retrieve digital content. These audio cues enable passive interactions, allowing users to receive relevant information directly on their smartphones without visual scanning or physical input. TISCODE leverages generative AI to synthesize distinct audio messages and employs a dual-stage decoding pipeline to ensure reliable classification across varying acoustic conditions. We evaluate the application prototype across multiple smartphones and test environments: indoor, outdoor, and noisy. Preliminary results are promising and support the system's suitability for real-world deployment. TISCODE represents a step toward inclusive, audio-driven interaction paradigms within the Internet of Audio Things (IoAuT), contributing to digital equity and accessible design.

CCS Concepts

• **Human-centered computing** → **Accessibility technologies**; • **Information systems** → *Multimedia information systems*; Test collections.

Keywords

Sound Processing, AI for Social Goodness, Sound Event Detection, Sound Classification, Internet of Audio Things.

ACM Reference Format:

Manuele Favero, Chiara Schiavo, Mirco Zavarise, Leonardo Badia, Sergio Canazza, and Luciano Murrone. 2025. Short Sound Messages for Enhancing Digital Accessibility and Reducing Digital Divide. In *International Conference on Information Technology for Social Good (GoodIT '25), September 3–5, 2025, Antwerp, Belgium.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3748699.3749813

*These authors contributed equally to this work.

GoodIT '25, Antwerp, Belgium

ACM ISBN 979-8-4007-2089-5/2025/09
https://doi.org/10.1145/3748699.3749813

1 Introduction

The development of digital technologies has significantly transformed the ways in which individuals access information, use services, and connect with each other. From communication platforms to smart infrastructures and ubiquitous computing, these innovations have increasingly become integral to daily life, enhancing productivity, connectivity, and convenience [39]. However, while the digital revolution has enabled unprecedented progress, it has also deepened existing structural inequalities, creating a digital divide between those who can fully benefit from technological advancements from those who remain marginalized. This divide is not merely a matter of access to devices or Internet connectivity, but extends deeply into the realm of accessibility, usability, and equitable participation in digital environments [17].

In this context, the role of technology as a force for social good has come under renewed scrutiny. Concepts such as Information and Communication Technologies (ICT) for goodness and AI fairness have emerged to emphasize the ethical responsibility of technologists and researchers to design systems that are inclusive by default [15, 36]. These principles urge us to go beyond abstract performance metrics and consider the societal impact of our designs, particularly in addressing the needs of vulnerable populations, including people with disabilities. Despite advancements in assistive technologies, many digital systems still fail to meet the diverse requirements of users with sensory, cognitive, or motor impairments [3]. The absence of inclusive design principles not only impairs accessibility but also reinforces systemic exclusions that manifest across education, employment, healthcare, and public life.

A particularly under-addressed subset of population consists of visually impaired or motion-impaired individuals, whose interactions with digital content are often mediated through assistive tools that are inconsistently supported or entirely absent in many contexts. Despite growing efforts to promote digital inclusion, individuals with disabilities struggle to gain access to Internet. According to [38], only 35% of people with disabilities reported using the Internet on daily basis, compared to 61% of those without impairments. Moreover, for 62.5% of individuals with disabilities, the cost of accessing ICT technologies is perceived as prohibitively high relative to their income, despite the potential of such technologies to significantly improve their daily lives [38]. Troubling statistics have estimated that, due to the lack of technologies that support independence, up to 30% of blind people do not travel independently outside their homes [10]. Even after accounting for income, education, and employment status, people with disabilities are, in fact, still less likely to access or use digital technologies, due to additional barriers such as the high cost of assistive tools, lack of accessible interfaces, and feelings of intimidation or low digital confidence. Psychological studies have documented the cognitive strain and emotional frustration experienced by these users when navigating environments that lack affordances for non-visual interaction [2, 19]. These studies further highlight the disconnect between current technological capabilities and the real-world needs of individuals who engage with digital systems differently from the normative user model.

An illustrative example of this disconnect is the widespread use of QR codes [34] and visually-oriented information displays in public and commercial spaces. While QR codes offer a compact and efficient mechanism for conveying digital content, their reliance on visual scanning and precise motor coordination renders them effectively inaccessible to many users with visual or motion impairments [9, 32]. The exclusion is particularly evident in environments such as transportation hubs, retail stores, and public service kiosks, where access to timely information can be critical and where no alternative modalities are offered [4, 27].

Recent works have explored the use of acoustic signals to encode digital content and enable short-range communication, highlighting their potential for accessibility [28] and wireless networking [18], though they remain untested in everyday applications.

In this paper, we present an experimental evaluation of a novel framework based on short sound messages (SSMs), named TISCODEs. These are brief sound cues designed to supplement or replace visual digital content in contexts where visual interaction is not feasible [7]. Each TISCODE, once generated, is paired with a digital information entry in the database and made available to the user, who can define when and where it should be transmitted. The TISCODE app operates passively, requiring no user interaction: when the corresponding SSM is emitted, the TISCODE app autonomously detects and decodes it, generating a pointer to the database in order to retrieve the associated content, which is then immediately made available on the user's smartphone. The linked digital information can include photos, videos, audio files, or any type of interactive multimedia content.

These SSMs can be transmitted and received not only by smartphones but also by a variety of IoT devices deployed in smart city environments. Such systems fall within the scope of the Internet of Audio Things (IoAuT) [37], a paradigm comprising devices capable of emitting and decoding acoustic messages.

Our goal is to mitigate accessibility barriers and reduce the cognitive and operational overhead typically required by assistive technologies, by embedding structured acoustic information into public interfaces. This approach is informed by recent research in human-computer interaction, accessible design, and auditory signal processing, and aligns with broader objectives of promoting ICT for social good and reducing the digital divide [15, 17].

Sound-based interaction is indeed often preferred because it enables eyes-free and hands-free communication, making it highly accessible for individuals with motor or visual impairments [21]. Unlike visual interfaces, auditory input and feedback can be used in dynamic or mobile contexts, enhancing usability and safety. Moreover, sound conveys emotional tone and urgency effectively, enriching user experience and expressiveness.

To support this vision, we discuss the design principles, implementation challenges, and potential of integrating such systems into everyday digital infrastructure. We present a performance evaluation of the current version of the TISCODE app, highlighting application scenarios where it shows promising results compared to traditional approaches relying on visual interfaces.

The remainder of this paper is organized as follows: Sec. 2 explores the various application scenarios of the TISCODE system, with a particular focus on its potential benefits for individuals with visual or motor impairments. Sec. 3 details the architecture of the TISCODE system. Sec. 4 outlines the experimental setup and methodology used for the performance evaluation. Sec. 5 presents and analyzes the experimental results. Finally, we conclude in Sec. 6.

2 TISCODE applications

The TISCODE system introduces a transformative approach to digital communication by leveraging the audio channel for the unidirectional transmission of encoded information. This method proves especially advantageous for individuals with visual or motor impairments, who often face significant barriers in interacting with conventional technologies such as QR codes, touchscreens, or location-based mobile notifications [24]. TISCODE eliminates the need for physical manipulation or visual engagement, enabling automatic, passive, and context-sensitive reception of information directly to the user's mobile device.

QR codes require users to visually locate and frame a printed code within a camera interface—a task that is impractical or impossible for blind or low-vision individuals. Similarly, motor-impaired users may find it difficult to hold a device, open a scanning application, and precisely align it with a code [26, 27]. TISCODE bypasses these requirements entirely: once the application is installed and configured with the user's preferences, any nearby TISCODE audio transmission is automatically decoded and presented on the smartphone. For blind users, this might mean receiving text-to-speech outputs, while motor-impaired users may benefit from automatic activation of assistive functions without any touch input.

This mechanism enables a diverse set of applications across public, commercial, and private domains:

Cultural Accessibility: In museums or cultural heritage sites, TISCODEs embedded in the background audio can trigger exhibit-specific content. A blind visitor approaching a sculpture may receive a notification with descriptive audio with synthesized speech describing the artwork and its context or high-contrast images (for low-vision users). Crucially, no physical interaction is needed: access to information is automatic, location-aware, and personalized.

Healthcare Environments: In hospitals or clinics, where navigation and orientation are often complex, TISCODEs can be used to broadcast room-specific information such as appointment details, check-in instructions, or route guidance. For a user with motor

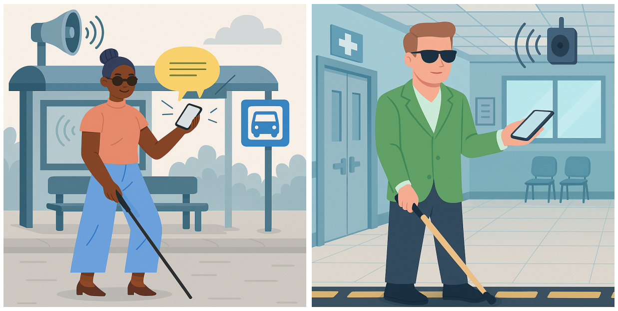

Figure 1: Examples of TISCODE applications: On the left, the use of TISCODE in an urban context—near a public transportation stop—to access real-time transit information. On the right, its deployment in a hospital setting to provide guidance on accessible routes within the facility.

disabilities, receiving these notifications without interacting with touchscreens simplifies the care experience [13, 21]. In emergency contexts, TISCODE-enabled public address systems can emit evacuation signals that instantly trigger multimedia safety instructions on users' smartphones, crucial for those who cannot rely on visual signage or alarms.

Retail and Commercial Spaces: In shopping centers or stores, localized TISCODEs can inform about offers, product highlights, or accessibility information (e.g., location of wheelchair-accessible facilities or elevators). A user with limited hand mobility may receive promotional content or navigation hints without opening any app or interacting with digital kiosks [27].

Urban Infrastructure: In recent years, the rising of smart cities has led to a growing deployment of IoT sensors in urban environments, collecting data related to traffic, pollution, and mobility [8]. These data, once processed, can be transmitted using TISCODE signals to enhance navigation and information access in outdoor settings, as shown in Figure 1. For instance, at bus stops or intersections, short jingle-like signals can trigger smartphone notifications that announce bus arrivals, delays, or safety warnings. This approach is particularly beneficial for visually impaired pedestrians, who may be unable to read posted schedules or safely interact with mobile interfaces while in motion [16]. With the support of TISCODE, which is broadcast through nearby transmission devices at the stop, users can receive schedule updates and interact with accessible services via voice commands on their smartphones.

Educational Campuses: On university campuses, TISCODEs can support students with disabilities by providing real-time updates and location-specific information. For instance, a student may receive class schedule changes or accessible route suggestions upon entering a building, triggered entirely by ambient audio cues.

Smart Homes and Assisted Living: In domestic environments, TISCODEs can be embedded in audio systems to automate reminders or alerts. A user with limited mobility may receive a multimedia notification with medication schedules or front-door activity whenever a TISCODE is emitted from a speaker or appliance. This passive signaling removes the need to operate complex interfaces or voice assistants [22, 35].

Overall, TISCODE explores the use of sound as a communication channel to improve accessibility in everyday environments. Its passive and non-intrusive approach aims to provide timely, location-aware information without requiring vision or fine motor skills. While still in development, the system represents a step toward more inclusive and accessible interactions within digital ecosystems.

3 TISCODE System

TISCODE is a novel patented technology [25] that leverages sound as a medium for short-range data communication in the IoAuT. A TISCODE is a sound message consisting in a fixed 1-second wake-up jingle, the Opening Marker (OM), and a 3-second INFOCORE, a sound that embed a pointer to digital content, which can be retrieved upon detection by a smartphone.

In this work, we use the term detection to refer to the identification of the OM, classification to indicate the matching of the INFOCORE with the sounds stored in the database, and recognition to denote the complete identification of the entire TISCODE, leading to the correct retrieval of the digital information associated with it.

The architecture of the TISCODE system comprises three key components: OM detection (Sec. 3.1), generative AI-based INFOCORE creation (Sec. 3.2), and INFOCORE classification and information retrieval (Sec. 3.3).

3.1 Opening Marker Detection

To enable passive operation, the TISCODE system includes an OM, a brief wake-up jingle designed to activate the recording and decoding processes on a smartphone. OM detection is performed using YAMNet [12, 23], fine-tuned to recognize the wake-up jingle class. The OM is designed to be perceptually distinct and spectrally disjoint from the payload, reducing false detections. Upon marker detection, the app begins recording and captures the ensuing INFOCORE segment for decoding.

This event-driven approach minimizes energy consumption and supports autonomous operation, making it suitable for real-time, context-aware applications in smart environments.

3.2 Generative AI for TISCODE

TISCODE system leverages generative AI to create distinct, information carrying INFOCOREs. As detailed in [7], it employs MusicGen [5], a transformer-based autoregressive model developed by Meta, to synthesize audio clips with high perceptual diversity. Each clip is generated from a structured prompt that defines musical attributes such as genre, instrument, tempo, key, and mood. This structured variability ensures that each audio jingle is unique and easily distinguishable [7].

To encode information within each INFOCORE, a total of 28 audio features are extracted. This set includes 25 low- and mid-level sound descriptors computed using traditional audio analysis tools such as Librosa [20]. The remaining three features are inferred using deep CNN-based models: CREPE [14] for key estimation and wav2vec [30] for genre and instrumentation. All features are subsequently quantized and mapped into binary bitmaps, producing a unique digital identifier for each INFOCORE.

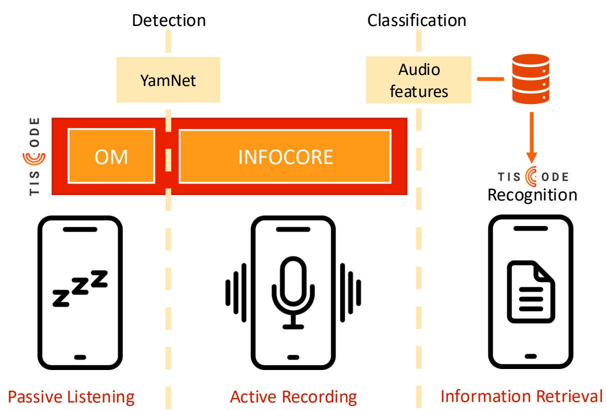

Figure 2: TISCODE system

Thanks to the use of generative AI and bitmap-based encoding, INFOCOREs can be dynamically synthesized on demand while ensuring a high degree of dissimilarity, both musically and computationally, across generated samples.

3.3 INFOCORE Classification

Once the TISCODE app detects the OM, it activates the microphone in active listening mode for three seconds to record the INFOCORE. The recorded clip is then sent to the server for decoding and classification.

During classification, the system re-extracts the same audio features used during INFOCORE generation and compares them to the stored bitmaps. Hamming distance minimization is applied for error correction [7]. In parallel, the classification algorithm performs frequency peak analysis via Short-Time Fourier Transform (STFT). It selects the top 10 frequency peaks per 0.5-second window and analyzes their temporal relationships. This spectral fingerprinting acts as a secondary verification mechanism to improve classification accuracy [11].

This dual-approach—bitmap decoding augmented by spectral analysis, ensures high recognition accuracy even in noisy environments and across different smartphone models with varying microphone responses [7].

4 Experiments

The TISCODE system has been analyzed and evaluated through a series of experiments conducted on five different mobile phones, each equipped with the TISCODE application. The goal is to validate the system's effectiveness as a real-time information retrieval solution based on SSMs.

The evaluation focuses on measuring the application's performances in accurately detecting and decoding the audio signals to retrieve the correct associated information. Furthermore, the experiments aim to identify the system's strengths, limitations, and potential bottlenecks, in order to guide future improvements.

This section is organized as follows: Subsec. 4.1 describes the experimental setup, including the devices, signal types, and evaluation metrics. Subsec. 4.2 presents a series of test scenarios and case studies, each designed to assess the system under different environmental and operational conditions.

Table 1: Specifications of the mobile devices used for TISCODE evaluation.

Device	Release Date	Release price
Samsung Galaxy A03	November 2021	159 €
Motorola Moto E7	November 2020	119 €
Oppo A54s	October 2021	229 €
Huawei Honor 9X	October 2019	269 €
Samsung Galaxy A55	March 2024	439 €

4.1 Set Up

The experimental setup consists of two fixed emitters and five receiving devices, used throughout all test sessions.

TISCODEs are reproduced using a 2025 MacBook Air 13", equipped with the Apple M4 chip. The receiving devices are five Android smartphones, each running the TISCODE application to record and decode the transmitted sound messages.

The selected smartphones represent a range of low- to mid-tier models, chosen to evaluate system performances across different hardware and software configurations. Specifically, the devices used are: Samsung Galaxy A03, Motorola Moto E7, Oppo A54s, Huawei Honor 9X, and Samsung Galaxy A55. Their release dates and prices are reported in Table 1. This diversity allows for a fair assessment of how the system performs with varying microphone quality, audio hardware, and processing capabilities, and supports the evaluation of the solution's robustness and scalability.

The sound signals used in this evaluation are taken from the dataset available in [6], filtered to include only TISCODEs with IDs from 101 to 200.

To ensure experimental consistency, all five smartphones are placed under identical conditions relative to the sound source, as shown in Figure 3. Each device is positioned on a flat, non-reflective surface, equidistant from the speaker, and oriented with the microphone facing the sound source. This setup is chosen to isolate the influence of hardware and signal processing characteristics on decoding performance, ensuring that differences in recognition accuracy are attributable to the devices themselves rather than to environmental variables.

The experimental setup is static, meaning neither the sound source nor the receiving devices are in motion during the tests. Initially, the phones are placed at a fixed distance of 0.5 meters from the speaker. In subsequent test sets, the distance is progressively increased to evaluate the system's range performance. Regarding audio levels, we consistently adopt four emission thresholds: 25%, 50%, 75%, and 100% of the MacBook Air's output volume. To quantify the acoustic intensity, we use the Sound Meter [29] application installed on each smartphone to measure the sound pressure level (SPL) at 0.5 meters. The SPL values are then averaged across all devices for each volume setting, resulting in approximately 44 dB, 47 dB, 53 dB, and 55 dB, respectively.

To simulate different acoustic environments, background noise and music were played through the second emitter, a Xiaomi Redmi

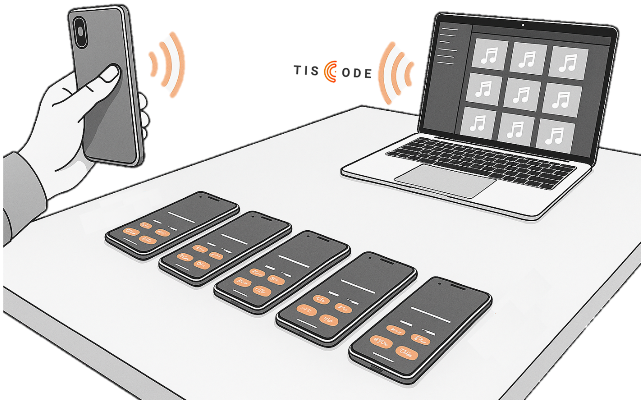

Figure 3: Experimental setup with a MacBook as the TISCODE emitter, five smartphones as receivers, and an additional smartphone generating background noise to simulate interference.

Note 9 Pro smartphone, positioned 0.3 meters above the receivers and at the same distance from the TISCODE source as the smartphones. The Xiaomi's volume was set to 40% for music and 60% for other ambient noise.

To ensure reproducibility and isolate the impact of environmental noise, we define a test configuration as a fixed combination of TISCODE playback volume and distance between the source and the smartphone. Each scenario corresponds to a specific background noise condition and is evaluated under one or more configurations.

For each configuration, we randomly select a set of 10 TISCODE signals, which are transmitted once per scenario. The same set is reused across scenarios sharing the same configuration, allowing for consistent comparison of system performance under different acoustic environments.

In its current version, the application can either detect the OM or fail. After the OM detection, the system proceeds to classify the INFOCORE, which can be correctly identified or result in a false positive. Detection results are computed as the number of detected OMs over the total number of emitted TISCODEs. Then, we calculate the correct classification rate as the ratio between the number of correctly classified (i.e. not result in a false positive) TISCODEs among those detected. If the OM of a TISCODE is not detected or the TISCODE is misclassified, the test is not repeated.

4.2 Test Scenarios

We structured the experimental evaluation into five scenarios: indoor and outdoor environments without artificial background noise; the third one includes the presence of background music; the fourth and fifth scenarios involve artificially added environmental sounds, simulating an urban setting and a hospital scenario, respectively. These controlled settings are designed to assess the robustness and reliability of the TISCODE system under different acoustic conditions, reflecting both ideal and real-world use cases.

The indoor environment corresponds to tests conducted in a quiet, enclosed room with negligible ambient noise. This setting enables precise measurement of system performance in near-ideal conditions and serves as a baseline reference. Conversely, the outdoor environment consists of tests carried out in a private residential garden, where natural and unpredictable background noises are occasionally present, such as wind, birds, traffic, pedestrians, and dogs. In these two contexts, background music will then be added, defining the third category, in order to assess the robustness of the application in the presence of musical interference.

The fourth and fifth scenarios introduce two real-world case studies designed to simulate environments where access to digital information is often limited for individuals with disabilities or special needs: the urban transportation scenario and the healthcare scenario.

Urban transportation scenario: In this scenario, we used background noise from urban transportation environments, such as traffic sounds, people conversing along the street [1] and atmospheric phenomena (rain, wind, thunderstorm) [33], to simulate a realistic context, such as a bus, metro, or tram stop. Visually impaired users often face difficulties accessing public transport schedules and real-time arrival information [16].

Healthcare scenario: Healthcare scenario mimics a hospital waiting area or entrance, where ambient noise is often composed of medical equipment beeps, announcements, and human voices [31]. In such contexts contactless and sound-based solution like TISCODE could deliver location-specific instructions or guidance without requiring visual interaction or manual input.

The goal of these experiments is to measure the detection and classification accuracy of TISCODE system. These findings may guide future deployment strategies, including the positioning of fixed TISCODE emitters in public spaces or institutions to maximize accessibility and reduce the impact of the digital divide for vulnerable user groups.

5 Results

This section presents the performance evaluation of the TISCODE application (April 2025 release), focusing on two key aspects: the detection accuracy of the OM and the classification accuracy of the INFOCORE. Tests were conducted across the described scenarios to assess the system's limitations, strengths, and robustness under varying acoustic conditions.

We first examine results from indoor and outdoor tests without background noise (Sec. 5.1), followed by experiments involving ambient music (Sec. 5.2). We then evaluate the system in two real-world scenarios, i.e. urban transportation and healthcare (Sec. 5.3), and conclude with a comparative analysis across all test conditions (Sec. 5.4).

5.1 Indoor and Outdoor Performances

Table 2 reports results from indoor tests, with the TISCODE emitted at a fixed distance of 0.5 meters from the smartphone microphone. When the output volume of the laptop is set to 75% or higher, all devices achieve 100% OM detection. At 50%, most devices maintain perfect accuracy, with the exception of the Galaxy A03, which drops to 90%. However, when the volume is reduced to 25%, performance degrades significantly. Both the Galaxy A03 and the Honor 9X fail to detect the TISCODE entirely (0% detection), while the Moto E7 exhibits only a slight drop in performance, reaching a 90% detection rate. Similar trends are observed in outdoor conditions (Table 5), where the lowest performance occurs again at 25% volume. Minor degradations are also observed at intermediate volume levels (50–75%), but detection remains above 50% when the volume is set to 50%.

When the volume is fixed and distance increases, as shown in Table 3 (indoor tests), we observe that, except for the Galaxy A03, all devices maintain a detection rate above 70% at 1.5 meters and never fall below 60% at 2 meters. Samsung devices tend to perform less reliably, with the Galaxy A55 being the only device to report false positives. In outdoor conditions (Table 6) performance generally declines slightly across all models, yet the Moto E7 consistently achieves OM detection rates above 70%, demonstrating superior robustness.

From these preliminary analyses, it is evident that the application yields satisfactory results, particularly at distances below 1.5 meters and with playback volumes above 50%. However, the overall performance is significantly influenced by the quality of the smartphone microphones, and the detection of the OM signal currently represents a bottleneck within the system.

5.2 Effects of Background Music

We evaluate recognition performances in the presence of ambient sounds and music. Table 4 presents a comparative analysis in an indoor setting, with and without background music, at a fixed distance of 0.5 meters from the sound source emitting the TISCODE at 50% volume. The results show that for the Moto E7 and the Oppo A54s, the only noticeable performance degradation occurs in the classification of the INFOCORE, likely due to interference from the music signal. In contrast, the Galaxy A03 and Honor 9X exhibit a significant drop in OM detection performance under the same conditions.

We can observe a similar, though slightly more pronounced, trend in the outdoor scenario with background music, as shown in Table 7. Despite the increased difficulty in detecting the OM, the overall classification accuracy remains comparable to the results obtained in the outdoor scenario without music, albeit with slightly reduced detection rates.

5.3 Urban and Healthcare Scenarios

We simulated two complex acoustic environments: an urban traffic scenario and an indoor hospital setting. For the urban environment, we consider multiple combinations of typical background noise sources, including vehicular traffic, weather-related sounds, and human speech.

Table 8 shows that the presence of speech significantly affects the detection of the OM, whereas traffic noise more critically impacts the classification accuracy of the INFOCORE signal. When analyzing the impact of traffic and weather noise at a distance of 0.5 meters, detection of the OM remains reliable even at 50% volume. However, at a distance of 1 meter, the TISCODE volume needs to be increased to at least 75% to improve classification performance, yet in only 2 out of 5 cases does it exceed a 60% classification rate.

When substituting traffic noise with human speech, the overall performance decreases further. Specifically, two devices, the Galaxy A03 and the Honor 9X, achieve only 40% and 30% recognition, respectively, at a 1-meter distance with the TISCODE volume set to 75%. In contrast, the remaining three devices consistently exceed a 60% recognition rate under the same conditions.

In the hospital scenario (Table 9), detection of the OM drops below 40% only when we set the volume to 50% at a 1-meter distance. In all other conditions, OM detection remains above 60%, with classification rates reached up a minimum of 89%.

5.4 Comparative Analysis

Figure 4 summarizes the overall system performances, averaging results across all smartphones and test conditions. Background music emerges as the most challenging scenario, particularly for OM detection. Since TISCODE is inherently designed as a short musical sequence, the presence of additional background music can interfere with accurate recognition. However, such a scenario is less likely to occur in real-world use cases, as TISCODE is intended for environments where loud background music is uncommon.

Across all conditions, average OM detection never falls below 54%, while INFOCORE classification remains above 82.75%. Notably, in Figure 5, the Moto E7 outperforms all other devices despite its low cost, with detection rates never below 85% and classification accuracy only dropping under background music conditions.

Figure 6 presents overall recognition performance in noisy environments at a 1-meter distance from the TISCODE source emitting at 50% volume. In this configuration, TISCODEs are correctly detected and classified in 72% of the cases. Focusing specifically on INFOCORE classification (Figure 7), the correct recognition rate increases to 90%, with only 10% of cases resulting in misclassification, as shown in Figure 7.

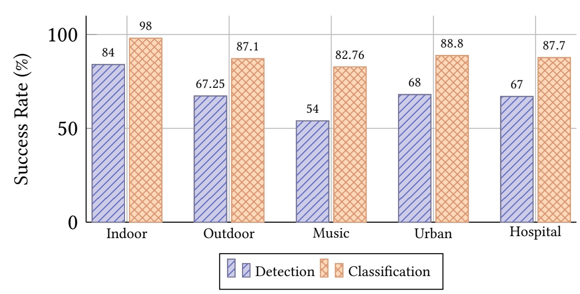

Figure 4: TISCODE APP comprehensive performances

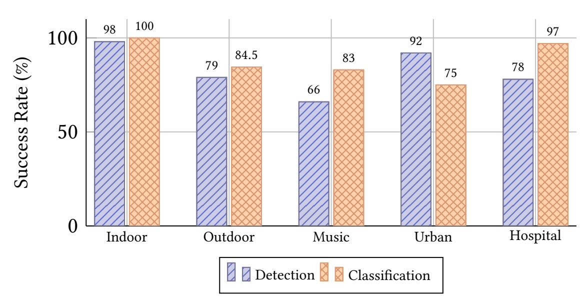

Figure 5: Moto E7 performances

Table 2: Indoor scenario: Average detection and correct classification rates for each device. The distance is fixed at 0.5m and the intensity varies (25%, 50%, 75%, 100%).

	Samsung Galaxy A03		Motorola Moto E7		Oppo A54s		Huawei Honor 9X		Samsung A55	
Intensity	Detection	Classification	Detection	Classification	Detection	Classification	Detection	Classification	Detection	Classification
25%	0%	0%	90%	100%	50%	100%	0%	0%	20%	100%
50%	90%	100%	100%	100%	100%	100%	100%	100%	100%	100%
75%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Table 3: Indoor scenario: Average detection and correct classification rates for each device. The intensity is fixed at 50% and the distance varies (0.5m, 1m, 1.5m, 2m).

	Samsung Galaxy A03		Motorola Moto E7		Oppo A54s		Huawei Honor 9X		Samsung A55	
Distance [m]	Detection	Classification	Detection	Classification	Detection	Classification	Detection	Classification	Detection	Classification
0.5	100%	100%	100%	100%	100%	100%	90%	100%	100%	100%
1.0	100%	100%	100%	100%	100%	100%	100%	90%	100%	100%
1.5	50%	100%	90%	100%	100%	100%	70%	100%	90%	100%
2.0	50%	100%	100%	100%	60%	100%	100%	100%	60%	66.66%

Table 4: Indoor scenario: Comparison between detection and correct classification rates with and without music under the same conditions. Distance is fixed at 0.5m and intensity at 50%.

	Samsung Galaxy A03		Motorola Moto E7		Oppo A54s		Huawei Honor 9X		Samsung A55	
Music	Detection	Classification	Detection	Classification	Detection	Classification	Detection	Classification	Detection	Classification
×	100%	100%	100%	100%	100%	100%	90%	100%	100%	100%
✓	50%	60%	100%	80%	100%	90%	10%	100%	70%	85.72%

Table 5: Outdoor scenario: Average detection and correct classification rates for each device. The distance is fixed at 0.5m and the intensity varies (25%, 50%, 75%, 100%).

	Samsung Galaxy A03		Motorola Moto E7		Oppo A54s		Huawei Honor 9X		Samsung A55	
Intensity	Detection	Classification	Detection	Classification	Detection	Classification	Detection	Classification	Detection	Classification
25%	0%	0%	40%	100%	70%	85.72%	20%	100%	10%	100%
50%	50%	100%	100%	80%	100%	100%	60%	83.33%	60%	83.33%
75%	90%	100%	100%	100%	100%	100%	100%	80%	100%	100%
100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Table 6: Outdoor scenario: Average detection and correct classification rates for each device. The intensity is fixed at 50% and the distance varies (0.5m, 1m, 1.5m, 2m).

	Samsung Galaxy A03		Motorola Moto E7		Oppo A54s		Huawei Honor 9X		Samsung A55	
Distance [m]	Detection	Classification	Detection	Classification	Detection	Classification	Detection	Classification	Detection	Classification
0.5	80%	75%	100%	90%	100%	80%	40%	75%	100%	90%
1.0	20%	100%	90%	77.77%	70%	85.71%	30%	33.33%	50%	100%
1.5	0%	0%	70%	71.42%	80%	50%	10%	100%	80%	50%
2.0	20%	100%	90%	77.77%	10%	0%	50%	60%	100%	80%

Table 7: Outdoor scenario: Comparison between detection and correct classification rates with and without music under the same conditions. Distance is fixed at 0.5m and intensity at 50%.

	Samsung Galaxy A03		Motorola Moto E7		Oppo A54s		Huawei Honor 9X		Samsung A55	
Music	Detection	Classification	Detection	Classification	Detection	Classification	Detection	Classification	Detection	Classification
×	80%	75%	100%	90%	100%	80%	40%	75%	100%	90%
✓	30%	66.66%	80%	75%	80%	87.5%	0%	0%	20%	100%

Table 8: Urban traffic scenario: comparison between detection and correct classification rates in an urban sound scenario with traffic, weather noises and background voices

					Samsung Galaxy A03		Motorola Moto E7		Oppo A54s		Huawei Honor 9X		Samsung A55	
Traffic	**Weather**	**Voices**	**Distance [m]**	**Intensity**	**Detection**	**Classification**	**Detection**	**Classification**	**Detection**	**Classification**	**Detection**	**Classification**	**Detection**	**Classification**
✓	✓	×	0.5	50%	100%	80%	100%	70%	90%	88.89%	70%	57,14%	100%	80%
✓	✓	×	0.5	75%	100%	90%	100%	90%	100%	100%	90%	77.77%	100%	100%
✓	✓	×	1	50%	0%	0%	80%	87.5%	10%	100%	30%	66.67%	60%	83.33%
✓	✓	×	1	75%	100%	80%	100%	70%	100%	90%	70%	57.14%	100%	80%
×	✓	✓	0.5	50%	100%	90%	100%	90%	90%	100%	10%	100%	70%	100%
×	✓	✓	0.5	75%	100%	100%	100%	100%	100%	90%	30%	100%	100%	100%
×	✓	✓	1	50%	0%	0%	0%	0%	10%	100%	0%	0%	0%	0%
×	✓	✓	1	75%	40%	100%	100%	90%	60%	100%	30%	100%	80%	100%

Table 9: Healthcare scenario: comparison between detection and correct classification rates in an hospital sound scenario.

		Samsung Galaxy A03		Motorola Moto E7		Oppo A54s		Huawei Honor 9X		Samsung A55	
Distance [m]	**Intensity**	**Detection**	**Classification**	**Detection**	**Classification**	**Detection**	**Classification**	**Detection**	**Classification**	**Detection**	**Classification**
0.5	50%	90%	100%	100%	90%	90%	88.89%	10%	100%	100%	100%
0.5	75%	100%	100%	100%	100%	90%	100%	30%	100%	100%	100%
1	50%	10%	100%	40%	75%	10%	0%	20%	50%	20%	50%
1	75%	60%	100%	100%	100%	100%	100%	70%	100%	100%	100%

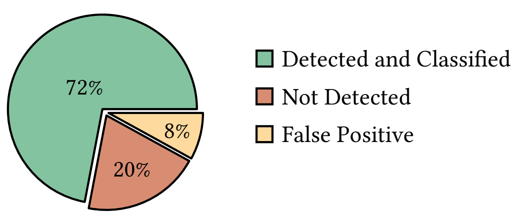

Figure 6: Average full recognition rates of TISCODE at a 1-meter distance and 50% volume under background noise conditions.

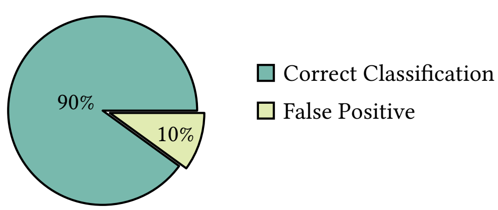

Figure 7: Average INFOCORE correct classification rates under various background noise conditions at a 1-meter distance and 50% playback volume.

6 Conclusion and Future Works

This work introduced and evaluated TISCODE, a sound-based system leveraging SSMs to facilitate equitable access to digital content, particularly for individuals with visual or motor impairments. Our experimental results, conducted across diverse environments and smartphone models, demonstrate that the TISCODE application can reliably classify sound messages and retrieve relevant digital content with considerable accuracy. By enabling passive, eyes-free, and hands-free interaction with digital information, TISCODE contributes to reducing cognitive and operational barriers often faced by users with disabilities. Its integration into public infrastructure and smart environments holds promise for bridging the digital divide and fostering accessible-by-design digital ecosystems. The application is still at the prototype stage and is not yet publicly available for download. All tests conducted so far have involved sighted individuals without disabilities. Once the application reaches a more mature stage, further evaluations will be necessary to assess the Quality of Experience for the target user groups, particularly those the system is intended to assist in reducing the digital divide.

Throughout the experiments, the main performance bottleneck identified was the detection of the OM, which consistently showed lower detection rates compared to the subsequent INFOCORE classification.

Future work will focus on improving the robustness of OM detection and enhancing the system's scalability to ensure practicality, adaptability, and real-world impact. Additionally, efforts will be directed toward defining a confidence threshold for INFOCORE classification. It is critical that the content delivered to the user corresponds exactly to the predefined information, especially in assistive contexts, to prevent confusion or distress. We will therefore explore a threshold mechanism that classifies uncertain or low-confidence TISCODE instances as "unrecognized" prioritizing safety and reliability over incorrect delivery—particularly in noisy environments.

By reshaping the way digital information is accessed and experienced, TISCODE paves the way for more inclusive, human-centered interaction paradigms, reinforcing the role of sound as a universal interface.

Acknowledgments

This work has been supported by the Italian PRIN project 2022PNRR "DIGIT4CIRCLE," project code P2022788KK.

References

[1] 2025. Freesound.org. [Online]. Available: https://freesound.org/ [Accessed: Apr 21, 2025].

[2] Forootan Amin. 2025. Empowering Agentic Non-Visual Web Navigation Through Tactile Controls and AI Support. (2025).

[3] Nicholas Caporusso, Lusine Mkrtchyan, and Leonardo Badia. 2009. A multimodal interface device for online board games designed for sight-impaired people. *IEEE Trans. Inf. Technol. Biomed.* 14, 2 (2009), 248–254.

[4] Abigail L Cochran. 2020. Impacts of COVID-19 on access to transportation for people with disabilities. *Transportation research interdisciplinary perspectives* 8 (2020), 100263.

[5] Jade Copet, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, and Alexandre Défossez. 2023. Simple and controllable music generation. *Adv. Neur. Inf. Proc. Syst.* 36 (2023), 47704–47720.

[6] Manuele Favero. 2024. TISCODE Dataset. (2024). https://www.kaggle.com/datasets/manuelefavero/tiscode-dataset

[7] Manuele Favero, Alessandro Buratto, Leonardo Badia, Sergio Canazza, and Luciano Murrone. 2025. Generative AI for Short Sound Message Transmission in the Internet of Things. In *Proc. IEEE Int. Conf. Machine Learning Commun. Netw. (ICMLCN)*.

[8] Manuele Favero, Chiara Schiavo, Alessandro Buratto, and Leonardo Badia. 2025. Price of Anarchy for Green Digital Twin Enabled Logistics. In *Proc. IEEE Symp. Comp. Commun. (ISCC)*.

[9] Mahmoud IM Fayyad, Maha M Sabar-Ayad, Abduelmula R Abduelkarem, et al. 2025. Medication management by visually impaired persons in the UAE: Self-Determination practice and law demands. *Pharmacy Practice* 23, 1 (2025), 1–10.

[10] Jenna L Gorlewicz, Jennifer L Tennison, Hari P Palani, and Nicholas A Giudice. 2018. The graphical access challenge for people with visual impairments: Positions and pathways forward. In *Interactive multimedia-multimedia production and digital storytelling*. IntechOpen.

[11] Jaap Haitsma and Ton Kalker. 2002. A highly robust audio fingerprinting system.. In *Proc. ISMIR*, Vol. 2002. 107–115.

[12] Andrew G Howard. 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861* (2017).

[13] Watthanasak Jeamwatthanachai, Mike Wald, and Gary Wills. 2019. Indoor navigation by blind people: Behaviors and challenges in unfamiliar spaces and buildings. *Brit. J. Visual Impairm.* 37, 2 (2019), 140–153.

[14] Jong Wook Kim, Justin Salamon, Peter Li, and Juan Pablo Bello. 2018. Crepe: A convolutional representation for pitch estimation. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 161–165.

[15] Irene Kitsara. 2022. Artificial intelligence and the digital divide: From an innovation perspective. In *Platforms and artificial intelligence: The next generation of competences*. Springer, 245–265.

[16] Wai-Ying Low, Mengqiu Cao, Jonas De Vos, and Robin Hickman. 2020. The journey experience of visually impaired people on public transport in London. *Transport Pol.* 97 (2020), 137–148.

[17] Stephen J Macdonald and John Clayton. 2017. Back to the future, disability and the digital divide. In *Disability and technology*. Routledge, 128–144.

[18] A. Madhavapeddy, R. Sharp, D. Scott, and A. Tse. 2005. Audio networking: the forgotten wireless technology. *IEEE Pervasive Computing* 4, 3 (2005), 55–60. doi:10.1109/MPRV.2005.50

[19] Tlamelo Makati, Garreth W Tigwell, and Kristen Shinohara. 2024. The Promise and Pitfalls of Web Accessibility Overlays for Blind and Low Vision Users. In *Proc. Int. ACM SIGACCESS Conf. Comp. Accessibility*. 1–12.

[20] Brian McFee, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. 2015. LIBROSA: Audio and music signal analysis in Python. In *Proc. Python Sci. Conf.* 18–24.

[21] Adib Mehrabi, Antonella Mazzoni, Daniel Jones, and Anthony Steed. 2020. Evaluating the user experience of acoustic data transmission: A study of sharing data between mobile devices using sound. *Pers. Ubiquitous Comput.* 24, 5 (2020), 655–668.

[22] Mohamed L Mekhalfi, Farid Melgani, Abdallah Zeggada, Francesco GB De Natale, Mohammed A-M Salem, and Alaa Khamis. 2016. Recovering the sight to blind people in indoor environments with smart technologies. *Expert Syst. Appl.* 46 (2016), 129–138.

[23] I Mohino-Herranz, J García-Gómez, S Alonso-Diaz, JG Gallegos, FJ Perez-Sanz, M Aguilar-Ortega, and R Gil-Pita. 2025. Implementing transfer learning for sound event classification using the realised audio database. *Measurement: Sensors* (2025), 101711.

[24] Aboubakar Mountapmbeme, Obianuju Okafor, and Stephanie Ludi. 2022. Addressing accessibility barriers in programming for people with visual impairments: A literature review. *ACM Trans. Accessible Comput. (TACCESS)* 15, 1 (2022), 1–26.

[25] Luciano Murrone. 2022. Method and system for transmitting information and data. US Patent Application US20230421516A1. Ogenus Srl, filed June 22, 2022.

[26] Hugo Nicolau, Tiago Guerreiro, Joaquim Jorge, and Daniel Gonçalves. 2014. Mobile touchscreen user interfaces: bridging the gap between motor-impaired and able-bodied users. *Universal access in the information society* 13 (2014), 303–313.

[27] Sunyoung Park, Hyun K Kim, Yuryeon Lee, and Jaehyun Park. 2023. Kiosk accessibility challenges faced by people with disabilities: an analysis of domestic and international accessibility laws/guidelines and user focus group interviews. *Univ. Access Inf. Soc.* (2023), 1–17.

[28] Xinghua Qu, Xiang Yin, Pengfei Wei, Lu Lu, and Zejun Ma. 2023. AudioQR: Deep Neural Audio Watermarks For QR Code.. In *IJCAI*. 6192–6200.

[29] rootApps. 2025. Sound Meter. [Online]. Available: https://play.google.com/store/apps/details?id=com.gamebasic.decibel [Accessed: Apr 21, 2025].

[30] Steffen Schneider, Alexei Baevski, Ronan Collobert, and Michael Auli. 2019. wav2vec: Unsupervised Pre-Training for Speech Recognition. In *Proc. Interspeech 2019*. 3465–3469.

[31] S. Shuvo. 2021. Hospital Noise Dataset. (2021). https://www.kaggle.com/datasets/nafin59/hospital-ambient-noise

[32] D. Sivaraj, Ram Sundar M, Sathish B, Suman M, and Madeshwaran S. 2024. Integrated Face Counting and Quick Response (QR) Object Detection with Voice-Based Alerts: A Smart Vision System. In *Proc. Int. Conf. Commun. Elec. Syst. (ICCES)*. 1608–1613. doi:10.1109/ICCES63552.2024.10860242

[33] J. H. Solorozano. 2022. Ambient Noise Dataset. (2022). https://www.kaggle.com/datasets/solor-zano/ambient-noise

[34] Tan Jin Soon. 2008. QR code. *synthesis J.* 2008, 3 (2008), 59–78.

[35] Dimitar H Stefanov, Zeungnam Bien, and Won-Chul Bang. 2004. The smart house for older persons and persons with physical disabilities: structure, technology arrangements, and perspectives. *IEEE Trans. Neur. Syst. Rehab. Eng.* 12, 2 (2004), 228–250.

[36] Nenad Tomašev, Julien Cornebise, Frank Hutter, Shakir Mohamed, Angela Picciariello, Bec Connelly, Danielle CM Belgrave, Daphne Ezer, Fanny Cachat van der Haert, Frank Mugisha, et al. 2020. AI for social good: unlocking the opportunity for positive impact. *Nature Commun.* 11, 1 (2020), 2468.

[37] Luca Turchet, György Fazekas, Mathieu Lagrange, Hossein S Ghadikolaei, and Carlo Fischione. 2020. The internet of audio things: State of the art, vision, and challenges. *IEEE Internet Things J.* 7, 10 (2020), 10233–10249.

[38] María Rosalía Vicente and Ana Jesús López. 2010. A multidimensional analysis of the disability digital divide: Some evidence for internet use. *The Information Society* 26, 1 (2010), 48–64.

[39] Kai Zhang, Xiaoting Cheng, Dan Li, and Xueling Meng. 2025. The Digital Divide of Older People in Communities: Urban-Rural, Gender, and Health Disparities and Inequities. *Health Soc. Care Communityy* 2025, 1 (2025), 1361214.