Remote Pointing over Telephony Channels: Audible-Sound Communication for Situated Industrial Support

Yuki Matsuda[1][0000−0002−3135−4915], Mayu Miyake[1], Taiga Tsuzuki[1], and Yoshihiro Kawakami[2]

[1] Okayama University, Okayama 700-0082, Japan
yukimat@okayama-u.ac.jp, mayu.miyake@cocolab.jp
[2] Kawakami K.K., Okayama 719-0241, Japan

Abstract. Remote troubleshooting in industrial maintenance remains challenging in environments with limited communication infrastructure. Many manufacturers continue to rely on conventional telephone calls to support equipment deployed at geographically remote sites with unreliable or unavailable internet connectivity. In such settings, video-based remote assistance cannot be used, and support interactions depend solely on verbal communication, which is often insufficient for equipment with complex structures and visually similar components. This paper proposes an audible-sound–based communication method over standard telephone lines that enables low-bandwidth data transmission alongside voice conversation. Digitally encoded information is superimposed onto audible sound signals that can pass through telephony audio channels, allowing the method to operate entirely within the audible frequency range without requiring additional network infrastructure. On the customer side, received sound signals are decoded by compact embedded devices and mapped to physical actuations, such as LEDs or buzzers installed on specific machine components. This enables remote operators to indicate precise physical locations on the customer's equipment during a phone call, transforming verbal instructions into situated, physical guidance. The proposed system incorporates dual-tone chords (two-frequency symbols) and parity-based error detection to ensure robustness under industrial noise conditions. A prototype implementation was evaluated in a real manufacturing environment using dried noodle binding machines. Experimental results demonstrate stable information transmission and reliable actuation even during machine operation. This work illustrates how legacy telephony infrastructure can support both data communication and ambient interaction, enabling effective remote industrial support in constrained environments.

Keywords: Customer support · Telephone support · Shared situational awareness · Audible-sound communication · Telephone lines.

1 Introduction

Many companies currently provide telephone-based support services to respond to customer inquiries and equipment-related issues or failures. Kawakami Co., Ltd., which develops and maintains dried noodle binding machines and weighing machines[3], also conducts failure response, troubleshooting, and confirming parts orders primarily through telephone communication. However, the equipment produced by the company has complex mechanical structures and contains many components with similar shapes and sizes. As a result, troubleshooting conducted solely via telephone suffers from limitations in verbal information sharing, making misunderstandings and rework likely.

Three issues frequently arise during telephone-based troubleshooting: (1) basic information sharing, (2) initial viewpoint sharing, and (3) target object location sharing. The basic information sharing issue occurs when the product model number used by the customer cannot be identified before troubleshooting begins. If the location where the model number is indicated is unclear, identifying the problem may take longer or lead to incorrect operational guidance. The initial viewpoint-sharing issue refers to the difficulty of determining from which position the customer is observing the equipment during a phone call. When operators cannot accurately understand the customer's viewpoint, they may provide incorrect instructions regarding direction or procedure. The target object location sharing issue occurs when either the operator or the customer fails to identify the instructed component correctly. This problem can result from unfamiliarity with component names, misidentification of model numbers, or discrepancies in initial viewpoint sharing that lead to an incorrect search area. These three issues are interrelated and collectively induce prolonged telephone support sessions. In addition, many noodle factories that use binding and weighing machines are located in mountainous areas where internet connectivity is often unavailable. Consequently, advanced communication methods such as video calls that rely on network infrastructure are difficult to use in these environments.

To address these challenges, this study aims to resolve the initial viewpoint sharing issue and the target object location sharing issue, which account for a significant portion of the time required in current telephone-based support. Using only existing telephone lines, we propose a technique that employs audible-sound–based communication to indicate specific locations on binding and weighing machines located at remote customer sites (Fig. 1). In the proposed system, operators transmit signal sounds that represent target locations, while the customer side extracts the signals from audio received through the telephone line and presents the corresponding locations using light or sound cues.

This paper describes the design of the audible-sound communication method and evaluates its effectiveness through preliminary experiments conducted within the manufacturing company. Experimental results confirm that the proposed method enables stable information transmission even in environments with significant noise, such as those generated by operating machinery.

[3] https://kawakami.tech/

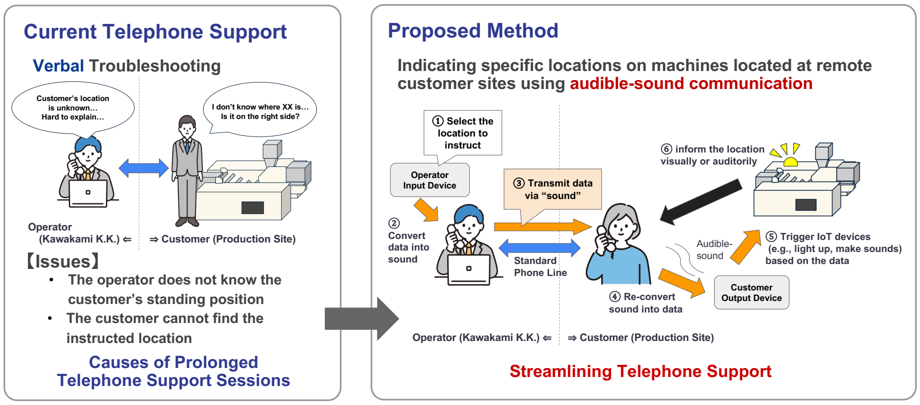

Fig. 1. Motivation and concept of this study

2 Related Work

2.1 Studies on Data Transmission Using Sound

The concept of transmitting encoded information over analog communication channels predates modern digital networks. One of the earliest examples is Bain's 1843 patent for the chemical telegraph, which introduced principles for scanning and synchronizing signals for remote reproduction [3]. Following the widespread adoption of voice telephony, Ives *et al.* further advanced this approach by demonstrating the transmission of high-fidelity images over standard telephone lines, laying the foundation for facsimile (fax) technologies [5]. These early milestones demonstrated that conventional audio channels can carry modulated non-verbal data, particularly for document reproduction.

In recent years, acoustic communication has attracted renewed attention as a means of ubiquitous data transmission. Many studies focus on *inaudible* frequency bands to embed information without disturbing human perception. Morise *et al.* proposed a method for embedding lyrics and artist information into music using mosquito sounds, enabling real-time extraction without perceptual degradation [7]. Suzuki *et al.* developed a technique for embedding video annotation data as high-frequency acoustic signals, facilitating video editing using smartphones [9]. Similarly, Ahn *et al.* enabled robot-to-robot communication without internet connectivity by employing inaudible frequencies [1], while Okano *et al.* achieved high robustness by combining signal superposition with information-theoretic redundancy, demonstrating recognition rates exceeding 99 % under noisy conditions [8]. Bai *et al.* proposed *BatComm*, a high-throughput inaudible acoustic communication system that leverages near-ultrasonic frequencies to enable data transmission between commodity devices [2]. Their work demonstrates that acoustic channels can support reliable and relatively high-rate communication, highlighting the potential of sound-based links as an alternative control and data channel for embedded and

IoT systems. Hanspach *et al.* demonstrated that low-bandwidth data transmission can be realized using acoustic signals propagating in air, forming covert acoustical mesh networks without relying on wireless or internet-based communication [4]. Their work highlights the feasibility of sound as a communication medium under constrained infrastructure conditions.

On the other hand, several studies have explored data transmission using *audible* frequency ranges. Katzschmann *et al.* demonstrated a system for controlling marine robots via underwater acoustic communication, emphasizing reliability in challenging environments [6]. Most relevant to the present study, Yang *et al.* proposed *DTMFTalk*, which utilizes Dual-Tone Multi-Frequency (DTMF) signals transmitted during voice calls for remote IoT control [10]. Their work demonstrates the effectiveness of audible sound communication for stable remote control in environments with limited internet connectivity.

2.2 Positioning of This Study

While the aforementioned inaudible communication methods offer the advantage of being imperceptible to humans, they are ill-suited for the legacy infrastructure addressed in this paper. Conventional telephone lines are designed primarily for voice communication, limiting the transmissible frequency bandwidth to a narrow range (typically 300 Hz–3.4 kHz). Consequently, high-frequency signals used in methods such as those by Suzuki *et al.* and Ahn *et al.* are cut off by the telephony system filter and cannot be reliably transmitted.

To address this limitation, this study adopts an audible-sound-based approach. Unlike conventional fax technologies aimed at static document reproduction, or IoT control via DTMF tones as in *DTMFTalk*, our method targets *situated physical actuation* (e.g., lighting up specific components) simultaneously with voice conversation. Although audible signals may be noticeable during a call, in a maintenance context, they serve a functional purpose: making the timing and presence of data transmission explicitly clear to users. This shared auditory awareness facilitates smoother collaboration between operators and customers, transforming the limitations of legacy audio channels into an interface for real-time guidance.

3 Proposed Method

3.1 Overview

Fig. 1 illustrates the overall concept of the proposed audible-sound-based information communication method over telephone lines. First, the operator selects the target location on the equipment using an input device. The selected information is then converted into sound signals and transmitted to the customer through an existing telephone line. On the customer side, the received sound signals are captured from the telephone handset and converted back into data. Finally, IoT devices installed on the equipment are activated according to the

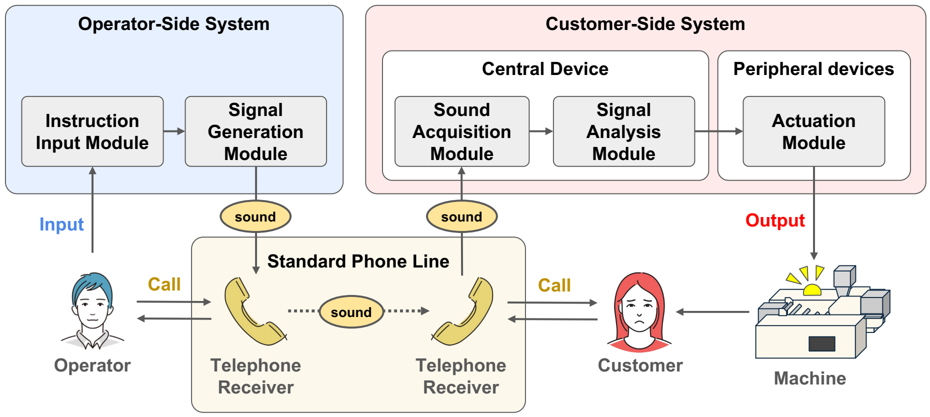

Fig. 2. Proposed system architecture

decoded data, such as emitting light or sound, to visually and audibly guide the customer to the location specified by the operator.

As shown in Fig. 2, the system consists of an operator-side system (instruction input module and signal generation module) and a customer-side system (sound acquisition module, signal analysis module, and actuation module). The customer-side system includes a central device that receives and decodes sound signals and peripheral devices that provide visual or auditory feedback. Each child device is attached to a specific location on the equipment that may need to be indicated during troubleshooting. The details of each component are described below.

3.2 Operator-Side System

The operator-side system runs on a PC and converts equipment operation instructions into audible sound signals for playback, as shown in Fig. 3. The system was developed using Python (Version 3.13.5) and PyQt6[4]. By selecting a target location (actuator ID) via the graphical user interface, the system generates and plays composite sound waves corresponding to the selected frequency components.

Instruction Input Module The instruction input module provides an interface that allows the operator to specify a particular target location on the equipment. During a phone conversation with the customer, the operator selects the actuator corresponding to the desired location using this interface, thereby determining the content of the transmitted signal. Although Fig. 3 shows an interface equipped with various configuration and monitoring functions for experimental purposes, the interface intended for actual operation is assumed to be a simplified version without these additional elements.

[4] `https://pypi.org/project/PyQt6/`

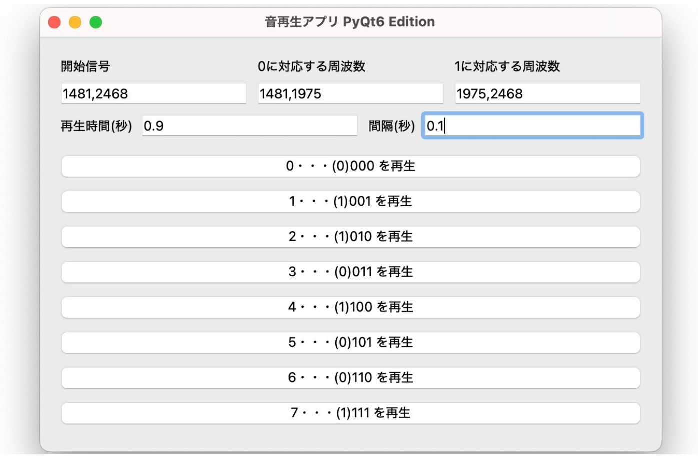

Fig. 3. Input interface of operator-side system (experimental version)

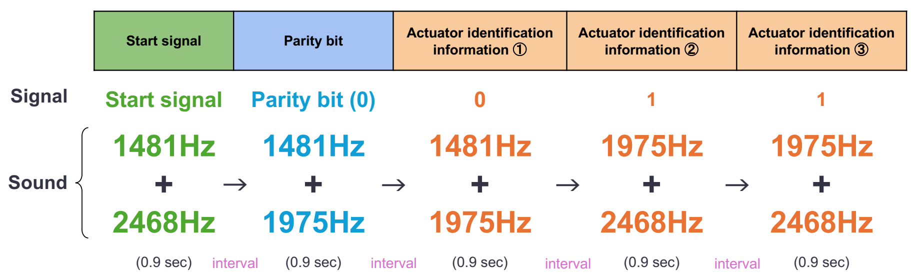

Fig. 4. Structure of the sound signal and an example (sending ID 3)

Signal Generation Module The signal generation module encodes the communication content determined by the instruction input module and generates sound signals for transmission over the telephone line. As shown in Fig. 4, each sound signal consists of a start signal for detecting the beginning of transmission, a parity bit for error detection (even parity in this study), and actuator identification information (3 bits in this study).

The signals are transmitted sequentially in the following order: start signal, parity bit, and actuator identification information. Each bit is represented by assigning two distinct frequencies, forming a chord. This design choice was made to reduce misrecognition caused by environmental noise, such as machine operation sounds commonly present in factories. The frequency combinations used for signal generation are listed in Table 1 and were selected to form harmonically acceptable chords that are not unpleasant to human listeners. Each sound signal has a fixed duration and interval (0.9 s and 0.1 s, respectively, in this study).

Table 1. Frequency combinations

Meaning	Frequency 1 (Note)	Frequency 2 (Note)
Start signal	1481 Hz (F#)	2468 Hz (D#)
0	1481 Hz (F#)	1975 Hz (H)
1	1975 Hz (H)	2468 Hz (D#)

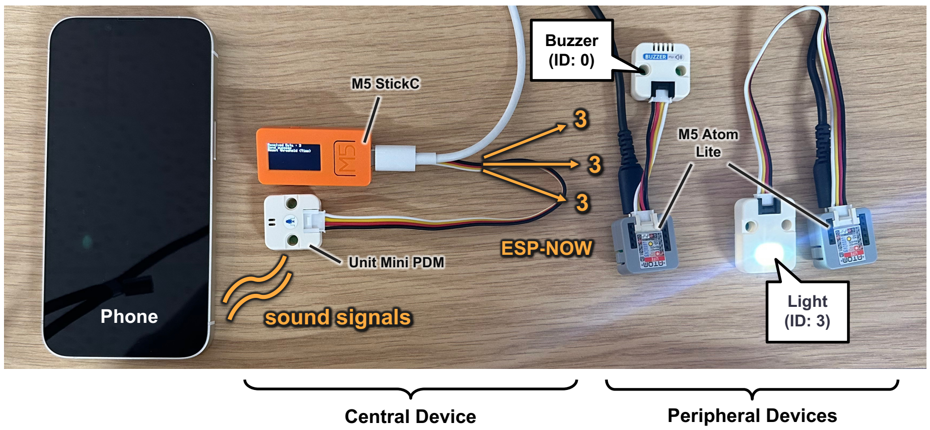

Fig. 5. Customer-side devices

3.3 Customer-Side System

The customer-side system consists of a central device that receives and analyzes sound signals transmitted over the telephone line and peripheral devices that present information to the user by being attached to the equipment, as shown in Fig. 5.

The central device uses an M5StickC[5] connected to an external microphone (Unit Mini PDM[6]) to receive and analyze sound signals and communicate with the peripheral devices. The Unit Mini PDM and M5StickC are connected via a GROVE cable, allowing real-time processing of captured audio data. Each child device uses an Atom-Lite[7] and is connected to actuators that indicate specific locations on the equipment. In this study, either a buzzer (Unit Buzzer[8]) or a light (Unit FlashLight[9]) is connected via a GROVE cable.

Sound Acquisition and Signal Analysis Modules (Central Device) The sound acquisition module captures audio from the customer-side telephone hand-

[5] `https://docs.m5stack.com/en/core/m5stickc`
[6] `https://docs.m5stack.com/en/unit/pdm`
[7] `https://docs.m5stack.com/en/core/ATOM%20Lite`
[8] `https://docs.m5stack.com/en/unit/buzzer`
[9] `https://docs.m5stack.com/en/unit/FlashLight`

set and performs frequency analysis. Fast Fourier Transform (FFT) FFT is performed at 10 Hz, and sound intensity is calculated at 250 Hz intervals from 250 Hz to 4000 Hz.

The signal analysis module extracts actuator identification information by separating and analyzing the sound signals. The procedure is as follows:

1. The start of a signal is detected by determining whether the sound intensity at frequencies corresponding to the start signal exceeds a threshold for 0.6 s.
2. After signal detection, reception of the parity bit and actuator identification information begins following a 0.4 s interval. Reception is confirmed when the corresponding frequencies exceed the threshold continuously for at least 0.7 s. If both 0 and 1 are detected due to environmental conditions, the value with the greater cumulative sound intensity is selected.
3. Signal reception continues for each 1 s symbol window until the expected signal duration is reached.
4. Error detection is performed using parity checking and verification of signal length.
5. If no errors are detected, the actuator identification information is converted from binary to decimal and broadcast to the actuation module using ESP-NOW[10].

The threshold is empirically set to three times the average sound intensity across the 250 Hz–4000 Hz range, based on preliminary experiments. To account for changes in environmental noise, the threshold is updated at three timings: immediately after system startup, after a fixed time interval (30 s in this study), and upon pressing a hardware button when no signal is being received.

The frequencies used for detection correspond to the nearest analyzed frequencies. For example, for the start signal assigned to 1481 Hz and 2468 Hz, detection is performed at 1500 Hz and 2500 Hz.

Actuation Module (Peripheral Devices) Each child device is assigned a unique ID and controls actuators installed at specific locations on the equipment. Peripheral devices continuously listen for broadcasts from the central device and determine whether the received ID matches their own. If a match is confirmed, the connected actuator is activated to indicate the target location specified by the operator.

4 Experiments

4.1 Experimental Overview

The experiments aim to evaluate the effectiveness of the proposed audible-sound communication method. We compare communication accuracy and received sound characteristics under different conditions, including telephone device combinations and the presence or absence of machine operation noise. Accuracy is evaluated based on whether the system successfully completes the process

[10] `https://www.espressif.com/ja-jp/solutions/low-power-solutions/esp-now`

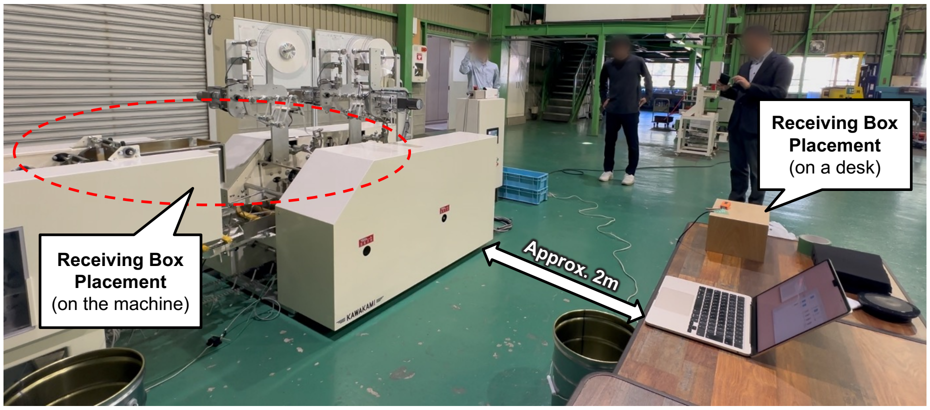

Fig. 6. Receiving box placed on a desk

of obtaining the correct ID in the signal analysis module. In addition, frequency analysis results obtained by the sound acquisition module are visualized for comparison.

4.2 Experimental Setup

A receiving box was constructed to control the placement of the customer-side telephone handset during experiments. A microphone was attached to the ceiling of the box, and the handset was placed inside with the screen facing upward and set to speaker mode. This setup was intended to reduce environmental noise on the customer side. The central device (M5StickC) was mounted on top of the box to facilitate threshold updates and monitoring. The operator-side handset volume was adjusted as needed.

To examine the effects of different telephone devices, experiments were conducted under two conditions: operator using a landline phone and customer using a smartphone, and both operator and customer using smartphones. To evaluate the impact of machine operation noise, three environmental conditions were considered: the receiving box placed on a desk with the machine inactive, on a desk with the machine active (Fig. 6), and on the machine itself with the machine active (Fig. 7). In total, six experimental patterns were evaluated. IDs from ID 0 to ID 7 were transmitted sequentially to verify whether 3-bit data could be reliably transmitted. Frequency analysis results were visualized for comparison.

The distance between the desk and the machine was approximately 2 m, and frequency analysis results were transmitted to a PC placed on the desk. The landline phone used was an ET-36Si-SDW (Hitachi), while the smartphones used were an iPhone 16e (Apple) on the operator side and an iPhone 14 (Apple) on the customer side.

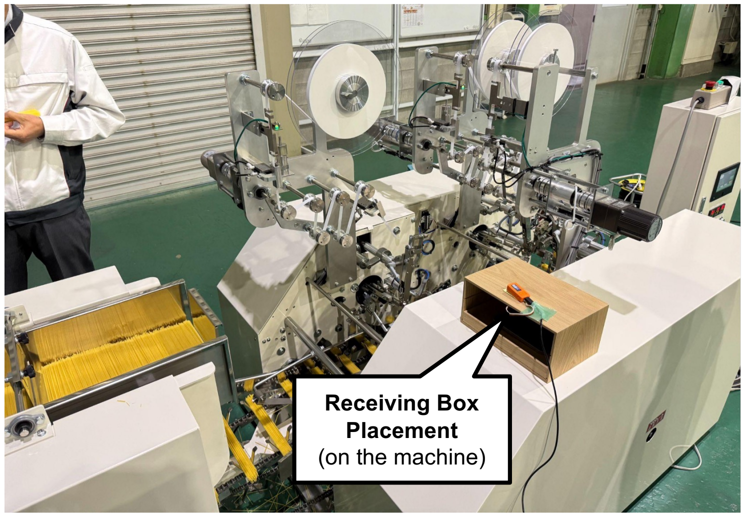

Fig. 7. Receiving box placed on the machine

4.3 Experimental Results

Landline Phone to Smartphone Fig. 8, Fig. 9, and Fig. 10 show the data reception results when the operator used a landline phone and the customer used a smartphone. In these graphs, the vertical axis represents sound intensity (microphone input), and the horizontal axis represents time. Frequencies used for signal reception (1500 Hz, 2000 Hz, and 2500 Hz) are highlighted in the figure (shown in green, orange, and light blue, respectively), while other frequencies are shown in gray. All 3-bit data were successfully received regardless of machine operation state or receiving box placement.

When the receiving box was placed on the desk and the machine was not operating (Fig. 8), the customer side had some background work noise; however, the signal sounds were clearly audible to human listeners even when they were away from the handset. Noise can be observed between ID 0 and ID 1, which was caused by conversation after transmitting the signal for ID 0. In contrast, when the machine was operating (Fig. 9), the operating noise and the signal sounds were perceived at approximately the same loudness level. Comparing Fig. 8 and Fig. 9, the amount of noise increases under machine operation; nevertheless, in both conditions, all transmitted data were correctly received. When the receiving box was placed on the machine while it was operating (Fig. 10), the operating noise and the signal sounds were also at comparable loudness. Although the noise was the largest among the three conditions, all transmitted data were successfully received.

Smartphone to Smartphone Fig. 11, Fig. 12, and Fig. 13 show the reception results when both the operator and the customer used smartphones. When the

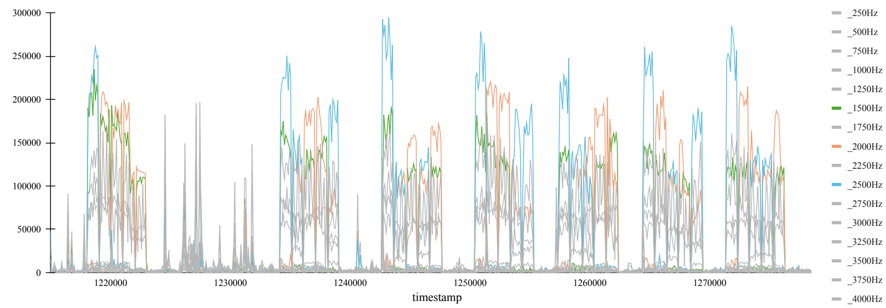

Fig. 8. Reception results (landline-to-smartphone / machine inactive / box on desk)

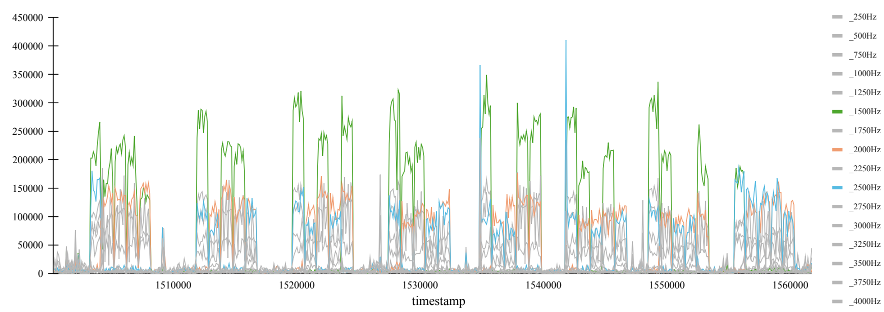

Fig. 9. Reception results (landline-to-smartphone / machine active / box on desk)

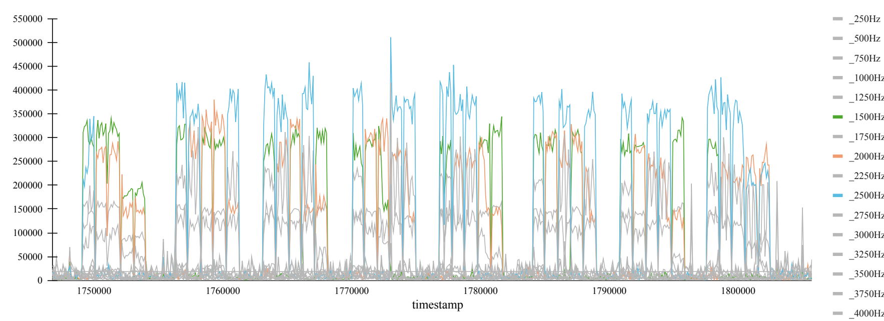

Fig. 10. Reception results (landline-to-smartphone / machine active / box on machine)

receiving box was placed on the desk, all 3-bit data were correctly received regardless of whether the machine was operating. However, when the receiving box was placed on the machine, reception failed for ID 2.

When the receiving box was placed on the desk and the machine was not operating (Fig. 11), the signal sounds were lower in volume than in the landline-to-smartphone condition, and there were moments when the customer was speaking. Even so, the signal sounds were clearly audible to human listeners even when they were away from the handset. When the machine was operating (Fig. 12), the operating noise was perceived as louder than the signal sounds. Although

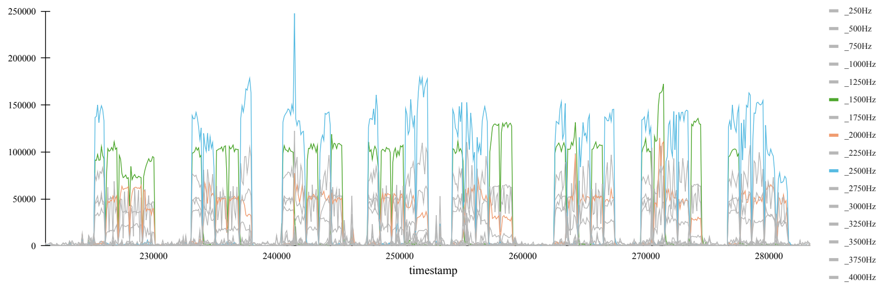

Fig. 11. Reception results (smartphone-to-smartphone / machine inactive / box on desk)

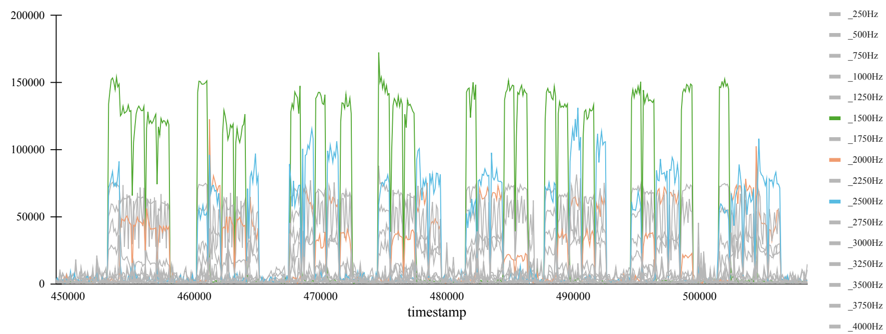

Fig. 12. Reception results (smartphone-to-smartphone / machine active / box on desk)

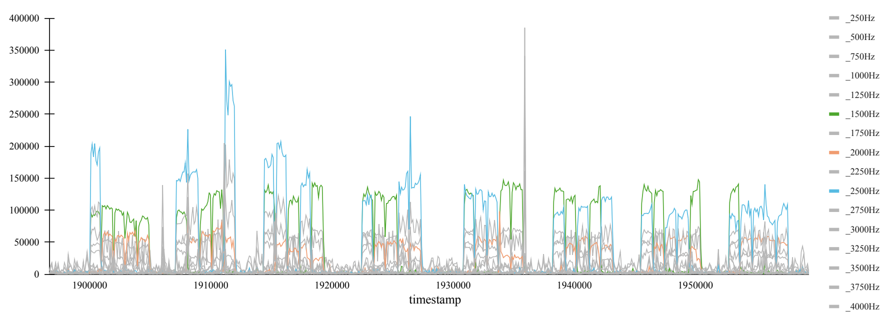

Fig. 13. Reception results (smartphone-to-smartphone / machine active / box on machine)

the noise level increased during operation as in the landline case, all transmitted data were still correctly received in this condition. In contrast, when the receiving box was placed on the machine while it was operating (Fig. 13), the operating noise was perceived as much louder than the signal sounds. In this condition, reception failed for ID 2. From the results, we confirmed that (i) the signal component around 2000 Hz became weaker and (ii) the placement made it easier to capture operating noise, leading to cases where the detection threshold exceeded the signal component. Nevertheless, since other IDs were successfully received, this issue is expected to be mitigated by introducing redundancy in transmitted data (e.g., repeated transmissions).

4.4 Discussion

The experimental results showed that all 3-bit data were correctly received in five out of six conditions (combinations of placement and telephone devices). This indicates that the proposed method can achieve stable information transmission even in environments with noise such as background work sounds, conversation, and machine operating noise. Moreover, the receiving box and the microphone attached to its ceiling enabled stronger reception of the signal sounds while suppressing surrounding noise to some extent.

From the graphs, we observed that sound intensity also tended to increase in the $\pm 250\,\mathrm{Hz}$ bands around the signal frequencies. Although this did not affect reception under the frequency combinations used in this study, care is required when selecting different frequency combinations.

In preliminary testing where the operator transmitted signals using a smartphone, we observed that the smartphone's noise-canceling function sometimes classified the signal sounds as noise and automatically attenuated them. In the present experiments, signal amplitudes in smartphone-based calls were generally lower than those in the landline condition, and this attenuation was particularly evident in the failed reception case. To address these issues, it is necessary to introduce redundancy (e.g., repeated transmissions) and to explore sound combinations that are less likely to be treated as noise by noise-canceling mechanisms.

The signal sounds used in this study were designed as harmonically acceptable chords, and they did not cause discomfort to human listeners. In addition, because the signal sounds are audible, the timing of transmission becomes explicit, which helps customers understand the system behavior during telephone support.

5 Conclusion

This paper aimed to address the initial viewpoint sharing issue and the target object location sharing issue, which account for a substantial portion of time in telephone-based support. Using only existing telephone lines, we designed an audible-sound communication technique that enables operators to indicate specific locations on remote equipment and evaluated its effectiveness through preliminary experiments conducted within the manufacturer.

The results demonstrated that 3-bit data were correctly received in five out of six conditions, indicating that stable information transmission is possible even in noisy environments, including those with machine operating sounds. On the other hand, the accuracy was found to depend on the loudness difference between environmental noise and the transmitted signal sounds. In addition, depending on the calling device and environment, there is a risk that noise-canceling functions may treat the signal as noise and attenuate it. To mitigate these issues, practical measures such as adding redundancy by transmitting signals multiple times are required.

As future work, we will investigate countermeasures for the issues identified in this study, evaluate cases where the receiver uses a landline phone, and conduct more rigorous experiments in environments closer to real-world deployment.

References

1. Ahn, S., Kim, D., Ahn, S.H.: Dynamic inaudible frequency shifting communication for multi-robot collaboration in manufacturing. Manufacturing Letters **44**, 1225–1231 (2025). https://doi.org/10.1016/j.mfglet.2025.06.142
2. Bai, Y., Liu, J., Lu, L., Yang, Y., Chen, Y., Yu, J.: BatComm: enabling inaudible acoustic communication with high-throughput for mobile devices. In: Proceedings of the 18th Conference on Embedded Networked Sensor Systems. pp. 205–217. SenSys '20 (2020). https://doi.org/10.1145/3384419.3430773
3. Bain, A.: Certain Improvements in Producing and Regulating Electric Currents and Improvements in Electric Time-Pieces and in Electric Printing and Signal Telegraphs. British Patent No. 9745 (1843)
4. Hanspach, M., Goetz, M.: On Covert Acoustical Mesh Networks in Air. arXiv preprint arXiv:1406.1213 (2014). https://doi.org/10.48550/arXiv.1406.1213
5. Ives, H.E., Horton, J.W., Parker, R.D., Clark, A.B.: The Transmission of Pictures over Telephone Lines. Bell System Technical Journal **6**(2), 187–214 (1927). https://doi.org/10.1002/j.1538-7305.1925.tb00946.x
6. Katzschmann, R.K., DelPreto, J., MacCurdy, R., Rus, D.: Exploration of underwater life with an acoustically controlled soft robotic fish. Science Robotics **3**(16), eaar3449 (2018). https://doi.org/10.1126/scirobotics.aar3449
7. Morise, M., Yamashita, Y.: Proposal on a lyric-synchronized music listening system using a watermark based on mosquito sound. IPSJ SIG Technical Reports **2012-MUS-96**(4), 1–6 (2012), `https://ipsj.ixsq.nii.ac.jp/records/83401`, (in Japanese)
8. Okano, Y., Teramoto, S., Noda, J.: Ad-hoc aerial acoustic communication for long-range propagation. In: IPSJ SIG Technical Reports, 2013-UBI-37(50). pp. 1–7 (2013), `https://ipsj.ixsq.nii.ac.jp/records/90985`, (in Japanese)
9. Suzuki, R., Sakamoto, D., Igarashi, T.: AnnoTone: Record-time Audio Watermarking for Context-aware Video Editing. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. pp. 57–66. CHI '15 (2015). https://doi.org/10.1145/2702123.2702358
10. Yang, S.R., Yuan, S.C., Lin, Y.C., Yang, I.F.: DTMFTalk: a DTMF-based realization of IoT remote control for smart-home elderly care. Mobile Networks and Applications **27**(1), 196–207 (2022). https://doi.org/10.1007/s11036-020-01641-0