



Development of a Protocol for Authenticating the Transmission of Login Credentials Between Devices Through Sound Tokens

Criação de um Protocolo para Autenticação da Transmissão de Credenciais de Login entre Dispositivos via Tokens Sonoros

Bruno Henrique Fonseca Stranak¹ , José Carlos Pizolato Junior² , Fabricio Archeleigar Kitazawa³

Received: March 15, 2026

Received in revised form: April 23, 2026

Accepted: May 13, 2026

Available online: May 19, 2026

ABSTRACT

This paper proposes an alternative authentication system for devices with limited input interfaces, eliminating manual typing or QR code scanning in the OAuth 2.0 protocol. The methodology establishes an acoustic communication channel in which a transmitter sends a verification code to a receiving smartphone. Frequency-Shift Keying modulation is used, operating between 15 and 20 kHz, a range silent to humans and less susceptible to common noise. Unlike established libraries such as *ggwave*, which use multi-tone modulation (MFSK), thus requiring high processing power and an audible frequency range to operate efficiently, the proposed methodology is structured for commercial equipment to operate at high frequencies with low computational cost. Decoding at the receiver uses an adaptation of the Goertzel Algorithm as a digital filter and envelope detector, ensuring low computational cost. Reverberation interference is also mitigated by temporal guard intervals. Tests varying distance, packet size, and transmission rate were analyzed via a Generalized Linear Model. Results indicate high robustness against noise, proving the viability of a functional and discrete authentication method for consumer devices.

keywords sound authentication, FSK modulation, scoustic, OAuth 2.0, Goertzel algorithm

RESUMO

Este trabalho propõe um sistema alternativo de autenticação para dispositivos com interfaces de entrada limitadas, eliminando a digitação manual ou o escaneamento de códigos por QR code no protocolo OAuth 2.0. A metodologia estabelece uma comunicação acústica onde um emissor transfere um código de verificação para um *smartphone* receptor. Utiliza-se a modulação por chaveamento de frequência (*Frequency-Shift Keying* – FSK), operando entre 15 e 20 kHz, região silenciosa aos humanos e menos suscetível a ruídos comuns. Diferente de bibliotecas consolidadas como a *ggwave*, que utilizam modulação multi-tom (MFSK), exigindo alto processamento e uma faixa audível de frequência para operar de forma eficiente, a metodologia proposta é estruturada para equipamentos comerciais operarem em altas frequências, com baixo uso computacional. A decodificação no receptor utiliza uma adaptação do Algoritmo de Goertzel como filtro digital e detector de envoltória, garantindo baixo custo computacional. A interferência por reverberação também é mitigada por intervalos de guarda temporais. Testes variando distância, tamanho do pacote e taxa de transmissão foram analisados via Modelo Linear Generalizado. Os resultados indicam alta robustez contra ruídos, comprovando a viabilidade de uma autenticação funcional e discreta em dispositivos de consumo.

palavras-chave autenticação sonora, modulação FSK, acústica, OAuth 2.0, algoritmo de Goertzel

¹Undergraduate student in Electrical Engineering, UFSCar, São Carlos, SP, Brazil. brunostranak@estudante.ufscar.br

²Prof. Dr., Department of Electrical Engineering, UFSCar São Carlos, SP, Brazil. jcpizolato@ufscar.br

³Undergraduate student in Statistics, USP, São Carlos, SP, Brazil. fa.kitazawa@usp.br

Introduction

The increasing interconnection of daily devices has brought new usability challenges, especially regarding user authentication. In devices with limited data entry, such as Smart TVs and game consoles, typing credentials with a remote control is a process that negatively affects usage efficiency and increases the likelihood of user errors, impacting the pillars of technology usability (Nielsen, 1993).

To overcome this, the market adopted the OAuth 2.0 Device Flow (Google Developers, 2025) and QR Codes. While these methods eliminate credential typing, they still require manual input or camera alignment, which can be inefficient and depend on the user's motor and visual conditions.

While the proposed solution aims to introduce an alternative method for authentication token validation, it may also have the secondary effect of reducing reliance on visual and motor interaction. This potential benefit, however, has not been specifically validated in real-world conditions. Therefore, although it could help address such accessibility-related constraints, further testing is required to assess the extent to which these advantages are actually achieved.

Considering efficiency in processing and the imperfect sound reproduction characteristics of consumer device speakers at high frequencies, this solution distances itself from state-of-the-art protocols like the ggwave library (Gerganov, 2022). Although such protocols represent the current standard in data transmission by sound, these consolidated solutions adopt multi-tone modulation (MFSK) and heavy error correction codes like Reed-Solomon (Tormim, 2023).

In practical scenarios involving consumer hardware, this approach may present limitations, as evidence suggests that intermodulation distortion at high frequencies can impact signal integrity (Tormim, 2023). While MFSK is a standard for high-density data, the use of serial FSK in this research aims to provide a lower computational overhead and greater resilience to the imperfect frequency response of common TV speakers. For this reason, the work proposes a leaner architecture using serial FSK modulation to ensure resilience against the frequency response of common TV speakers. Decoding is performed via a modified Goertzel Algorithm, Puhlmann (2015), acting as a resonant filter and envelope detector to minimize computational overhead and mitigate reverberation through guard intervals.

Data transmission via sound signals utilizes the hardware infrastructure already present in most electronic devices: microphones and speakers. The proposal uses physical space as a channel to transmit an authentication token via sound signals. The performance evaluation of the proposed system was conducted in environments with realistic acoustic characteristics and under usual practical conditions. It is important to emphasize that this work focuses on the development and validation of the token transport layer, namely, communication through an acoustic environment.

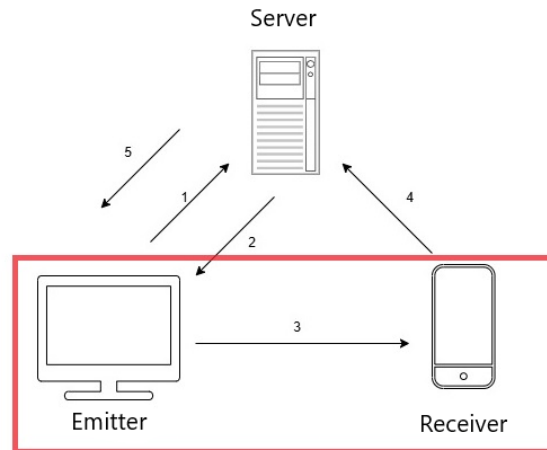
Material and methods

System architecture and communication protocol

Communication between devices was designed to integrate with a market-consolidated authentication system, the OAuth 2.0 Device Authorization Grant, officially defined by RFC 8628. This protocol was specifically developed for scenarios involving devices with data entry limitations, such as Smart TVs and video game consoles, where typing credentials is done via a virtual keyboard in which the user must select one letter at a time, demanding time and patience.

The system proposed in this work acts on a specific and crucial step of this flow: the transfer of the verification code from the primary device (the Emitter, e.g., the TV) to the user's secondary device (the Receiver, e.g., the smartphone). The protocol does not specify how this data transfer process should be established; however, it is currently performed manually by the user (either by typing the code or scanning a QR Code). Our approach replaces this interaction with a transmission of sound signals in an acoustic environment.

Figure 1 illustrates the diagram based on the architecture documented in Denniss et al. (2019), showing the complete authentication path and highlighting in red the step that constitutes the scope of this work.

Figure 1 - Sequence diagram of authentication steps using OAuth architecture.

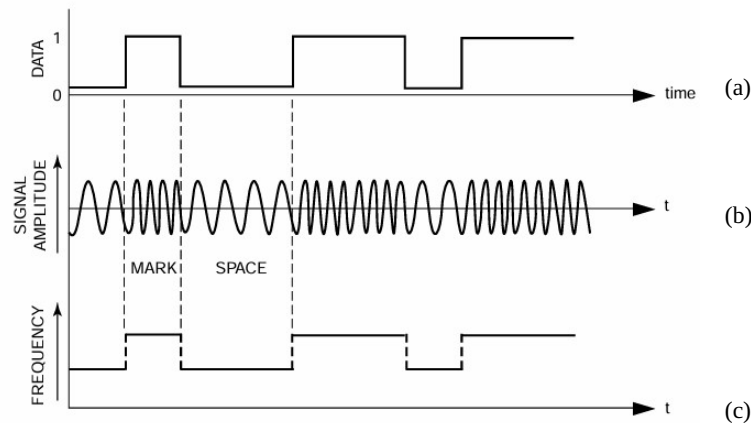
The authentication steps are described as follows:

1. The limited-input device (e.g., Smart TV) initiates the login process and requests a user access code from the authentication server.
2. The server generates a code and sends it back to the emitter; additionally, the verification site address is also sent. However, in this work, due to character limitations during transmission, this address must be previously saved on the receiver side. Traditionally, the token is displayed on the emitter's screen for the user to see.
3. This is the central step and the main contribution of the research. Instead of the user needing to manually type the code or scan a QR Code with a smartphone (receiver), an alternative sound-based approach is proposed.
 - (a) The emitter encodes this access code/token into a sound signal using Frequency-Shift Keying (FSK) modulation and emits the sound through the speakers.
 - (b) This acoustic signal is transmitted to the Receiver, which uses its microphone to "listen" to the sound sequence.
 - (c) The receiver then decodes the sound to recover the original token, using the Goertzel Algorithm to detect frequencies.
4. With the token received acoustically, the application on the receiver sends this code to the server to authorize access on behalf of the user, who is already logged into the receiver's application, completing verification on the secondary device.
5. After the Receiver authorizes access, the server sends the final authentication token to the emitter, which finishes the login process and grants access to the content.

Signal encoding and transmission —————

To enable step 3 of the flow presented in Figure 1, where the token is transferred over the air, it was necessary to define a modulation method robust enough to operate on common TV speakers. Among digital options, FSK was chosen, as described below, due to its simplicity and resistance to amplitude variations, which are essential characteristics for the acoustic channel (Watson, 2001).

Frequency-Shift Keying (FSK) is a digital modulation technique used to transmit binary information (Watson, 2001). The fundamental principle of FSK consists of representing digital data ('0' and '1') or ('space' and 'mark') by varying the frequency of a continuous carrier wave. For each bit of information to be transmitted, the signal frequency is switched to one of two predefined frequencies (Watson, 2001). Figure 2 shows the FSK modulation process.

Figure 2 - FSK Modulation: (a) Binary data; (b) Modulated signal; (c) Signal frequency characteristic.

From "FSK: Signals and Demodulation" by B. Watson, 2001, WJ Communications, Inc.

Signal decoding based on Goertzel Algorithm

Once the signal is FSK-modulated by the emitter, the challenge on the receiver side (smartphone) lies in the efficient detection of these frequencies in real-time. Considering the limited processing capacity of mobile devices, implementing a digital matched filter becomes the ideal approach.

Optimal detection or demodulation of FSK signals in noisy channels is traditionally achieved with two receiver architectures that, while theoretically equivalent, differ in their practical approach: the correlation receiver and the matched filter receiver (Guimarães & de Souza, 2015).

The correlation receiver operates in the time domain. Its methodology consists of internally generating reference sine and cosine waves, which are also perfectly "matched" with the mark and space frequencies. The received signal is multiplied by these references, and the results are integrated (summed) over the bit duration. The final decision is based on the total accumulated energy; the reference frequency resulting in the highest final sum determines the decoded bit (Guimarães & de Souza, 2015).

The matched filter receiver methodology acts as a filtering process. Like the correlator, its filters are "matched" to the frequencies of interest, but signal processing occurs differently. The input signal passes through a bandpass filter (the matched filter), which transforms it into a new sampled signal, a high-frequency pulse whose amplitude represents the presence of the desired tone (Watson, 2001). The crucial difference lies in the treatment of this output signal: instead of a single final sum, the pulse is processed by an envelope detector, which performs rectification followed by low-pass filtering over all samples of the pulse to extract a smoothed energy envelope (Guimarães & de Souza, 2015).

While both architectures are fundamentally "matched" to target frequencies, the practical implementation of the matched filter receiver, with its additional low-pass filtering stage on the envelope, can achieve superior performance. Guimarães and de Souza (2015) demonstrate through digital simulations that this additional filter provides a "reduction in noise variance," making the final energy measurement more robust.

In this work, the Goertzel Algorithm will be used as the digital implementation of the matched filter methodology. This choice is justified because the algorithm is a specific form of the Discrete Fourier Transform (DFT), aimed at searching for specific frequencies rather than a broad spectrum, making signal processing efficient (Puhmann, 2015). The implementation will be modified to replicate the high-robustness architecture described by Guimarães and de Souza (2015): instead of using only the final energy value of the algorithm, intermediate states of each sample within the chunk will be considered. On this sequence of values, rectification and low-pass filtering via a moving average will be applied to smooth the energy envelope and dilute noise peaks to obtain a more reliable final decision.

The detection of specific frequencies in a digital signal is a fundamental task in communication systems. The Discrete Fourier Transform (DFT) is the mathematical tool that allows analysis of the frequency spectrum of a sampled signal (Bouchaleun, 2019). However, for applications where the interest lies in a small number of known frequencies, such as FSK demodulation, calculating the full spectrum through algorithms like FFT becomes computationally inefficient (Puhmann, 2015).

In this context, the Goertzel Algorithm presents itself as an optimized particularization of the DFT. The main advantage of the algorithm is its ability to calculate the spectral content of a single frequency ("bin") at a time rather than the full spectrum, which considerably reduces the volume of calculations required for detection. This efficiency makes it ideal for practical applications, such as DTMF (Dual-Tone Multi-Frequency) tone detection in telephony systems, where the algorithm replaced old analog filters (Puhlmann, 2015).

The Goertzel Algorithm operates in two phases on a block of N samples of an input signal $x(n)$. The first phase is a recursive process using a second-order IIR (Infinite Impulse Response) filter, and the second consists of the final calculation resulting in the desired spectral component (Puhlmann, 2015).

First Phase (Recursive) In this stage, the N signal samples are processed sequentially. This is evident in equations (1) – (3):

$$z(0) = \text{coefk} \cdot z(1) - z(2) + x(n); \quad (1)$$

$$z(2) = z(1); \quad (2)$$

$$z(1) = z(0), \quad (3)$$

where $x(n)$ is the current sample of the input signal, $z(2)$ corresponds to $z(0)$ delayed by two samples, $z(1)$ corresponds to $z(0)$ delayed by one sample, and $z(0)$ represents the accumulated result of the first phase. The parameter coefk is a pre-calculated coefficient that tunes the algorithm to the frequency of interest, defined by equation (4):

$$\text{coefk} = 2 \cdot \cos(2 \cdot \pi \cdot k/N). \quad (4)$$

The term k represents the frequency index to be detected and is determined by the relationship between the frequency of interest (F_i), the sampling frequency (F_s), and the number of samples (N), given by $k = N \cdot F_i/F_s$ (Puhlmann, 2015).

Second Phase (Final Calculation) After the recursive loop ends, a final calculation determines the real and imaginary components. The magnitude, representing the signal energy, is then obtained from these components (Puhlmann, 2015), as defined by equations (5) and (6):

$$y_{\text{real}} = \frac{z(1) - z(2) \cdot \cos(2 \cdot \pi \cdot k/N)}{N}, \quad (5)$$

$$y_{\text{img}} = \frac{z(2) \cdot \sin(2 \cdot \pi \cdot k/N)}{N}. \quad (6)$$

Finally, the squared magnitude, proportional to energy and sufficient for comparison, is calculated in equation (7):

$$y_{\text{mag}}^2 = y_{\text{real}}^2 + y_{\text{img}}^2. \quad (7)$$

As established, the matched filter receiver architecture followed by an envelope detector demonstrated superior robustness due to its ability to reduce noise variance in the decision variable (Guimarães & de Souza, 2015). To digitally replicate this behavior, the Goertzel algorithm was substantially adapted.

The classic Goertzel implementation iterates recursively over the N samples and only at the end calculates a single pair of components. In contrast, in this work, the algorithm was modified to operate as a discrete-time resonant IIR filter. Thus, intermediate outputs of the recursive loop are interpreted as a time-domain filtered signal, which is rectified and subjected to a moving average filter, resulting in the amplitude envelope estimate. This adaptation allows monitoring the energy evolution over time, increasing robustness against noise and rapid symbol transitions.

System technical parameters and reproducibility —

To address the requirement for full system reproducibility, this section details the exact numerical parameters used in the development of the acoustic authentication protocol. The system was implemented using a sampling rate (F_s) of 48,000 Hz, synchronized between the Python-based emitter and the Kotlin-based receiver.

Table 1 summarizes the physical and digital signal processing (DSP) constants. In Algorithm 1, the implementation of the modified Goertzel algorithm is detailed in pseudocode. In Algorithm 2, the resonant time-series output from Algorithm 1 is rectified and smoothed via a 100-sample moving average filter to dilute impulsive noise and extract a stable amplitude envelope for the final decision logic.

Table 1 - Digital signal processing parameters.

Parameter	Symbol / Variable	Value
Sampling Rate	F_s	48,000 Hz
Frequency of interest	F_{target}	15, 18 or 20 kHz
Bitrate	-	-
Bit 0 Frequency	F_0	15,000 Hz
Bit 1 Frequency	F_1	18,000 Hz
Start/Sync Frequency	F_{start}	20,000 Hz
Symbol Duration	T_{sym}	1/Bitrate ms
Guard Interval	T_{guard}	$2 * T_{sym}$ ms
Goertzel Block Size	N	1,600 samples
Moving Average Window	W	100 samples
Amplitude Threshold	-	2,500.0
Fade Ramp Duration	-	5 ms
Alpha Numeric Token	-	randomized

The core technical enhancement of the receiver lies in the transition from a static magnitude estimator to a continuous resonant filter. The differences and the resulting benefits for signal precision are detailed below:

- **From Static Magnitude to Time-Series:** The classical Goertzel implementation is typically used as an integrator that yields a single energy value for an entire block of samples. In contrast, our adaptation utilizes the recursive IIR stage to capture all intermediate state variables ($q_0[n]$). This transforms the algorithm into a filter that generates a resonant time-series, preserving the temporal evolution of the target frequency.
- **Peak Dilution via Moving Average:** In the classical approach, a high-energy impulsive noise event (such as a mechanical click) occurring within the sample block would be integrated into the final result, artificially inflating the magnitude and potentially causing false detections. By extracting intermediate values, our system allows the application of a moving average filter across the rectified series.
- **Enhanced Detection Precision:** This combination ensures that transient noise peaks are "diluted" across the integration window. Consequently, the system requires a sustained presence of the carrier frequency to register a high amplitude, rather than a single instantaneous burst of energy. This significantly increases the precision of the receiver in real-world acoustic environments.

Algorithm 1 Modified Goertzel implementation as a time-series filter

```

1: procedure GoertzelAsFilter(Samples,  $F_{target}$ ,  $F_s$ )
2:    $N \leftarrow \text{Samples.size}$ 
3:    $k \leftarrow \lfloor 0.5 + \frac{N \cdot F_{target}}{F_s} \rfloor$  ▷ Target frequency index calculation
4:    $\omega \leftarrow \frac{2\pi k}{N}$ 
5:    $coeff \leftarrow 2 \cdot \cos(\omega)$  ▷ Filter feedback coefficient
6:    $q_1, q_2 \leftarrow 0, 0$  ▷ State variables (delay elements)
7:   FilteredSignal  $\leftarrow$  Array of size  $N$ 
8:   for  $n = 0$  to  $N - 1$  do ▷ Recursive IIR stage
9:      $q_0 \leftarrow coeff \cdot q_1 - q_2 + \text{Samples}[n]$ 
10:    FilteredSignal[ $n$ ]  $\leftarrow q_0$  ▷ Extracting the resonant time-series
11:     $q_2 \leftarrow q_1$ 
12:     $q_1 \leftarrow q_0$ 
13:  end for
14:  return FilteredSignal
15: end procedure

```

Algorithm 2 Envelope extraction with moving average smoothing

```

1: procedure GetEnvelope( $Samples, F_{target}, F_s$ )
2:    $FilteredSeries \leftarrow$  GoertzelAsFilter( $Samples, F_{target}, F_s$ )
3:    $RectifiedSeries \leftarrow$  Array of size  $N$ 
4:    $SmoothedEnvelope \leftarrow$  Array of size  $N$ 
5:    $W \leftarrow 100$  ▷ Moving Average window size
6:   for  $n = 0$  to  $N - 1$  do
7:      $RectifiedSeries[n] \leftarrow |FilteredSeries[n]|$  ▷ Full-wave rectification
8:   end for
9:   for  $n = W$  to  $N - 1$  do ▷ Applying Moving Average
10:     $Sum \leftarrow \sum_{i=n-W+1}^n RectifiedSeries[i]$ 
11:     $SmoothedEnvelope[n] \leftarrow \frac{Sum}{W}$  ▷ Peak dilution and smoothing
12:  end for
13:  return  $\max(SmoothedEnvelope)$  ▷ Returns the stable peak amplitude
14: end procedure

```

Other acoustic interference mitigation strategies —

While the aforementioned decoder efficiently handles digital signal processing, the physical communication channel, in a closed environment introduces distortions like noise and reverberation (echo). To ensure signal quality, complementary acoustic strategies were adopted at the physical layer.

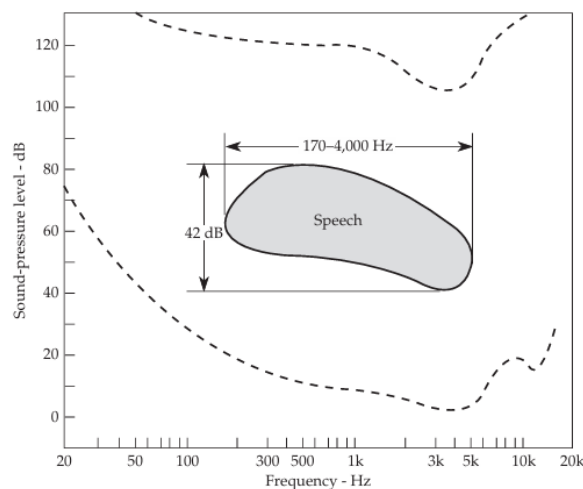
Acoustic communication performance in closed rooms is governed by background noise and reverberation. As Bistafa (2018) points out, the Signal-to-Noise Ratio (SNR) is critical for intelligibility. Simultaneously, reverberation, the persistence of sound after the source stops (Everest & Pohlmann, 2021), is the main cause of Intersymbol Interference (ISI), where previous bit "echoes" overlap current bits (Proakis & Salehi, 2008). Consequently, this project adopted two strategies: (1) Strategic operating frequency range; (2) Treatment of reverberation and ISI.

Operating frequency range selection —

The choice of frequencies determines system robustness and usability. Data channels (F0 and F1) and control (START) were allocated between 15 and 20 kHz based on two premises: seeking spectral regions with lower noise density and minimizing user perception of the transmission.

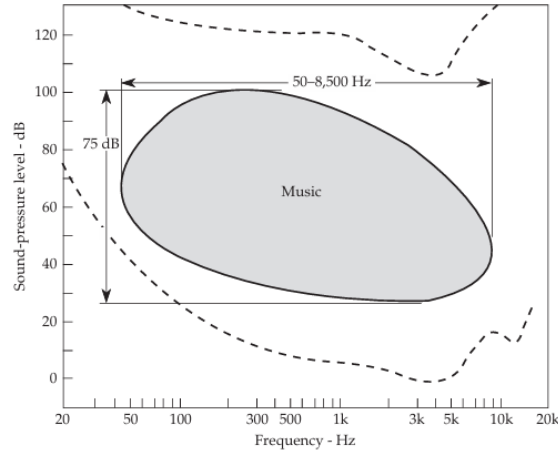
Acoustics literature shows that common residential sounds concentrate energy in central auditory ranges. Closed environment analysis by Vieira and Costa (2012) shows relevant sound energy between 125 Hz and 4,000 Hz. Specifically, human speech, Figure 3, and music, Figure 4, mostly occupy regions below 8,500 Hz (Everest & Pohlmann, 2021).

Figure 3 - Auditory region used for speech.



From "Master handbook of acoustics" by F. A. Everest & K. C. Pohlmann, 2009, McGraw-Hill Education.

Figure 4 - Auditory region used for music.

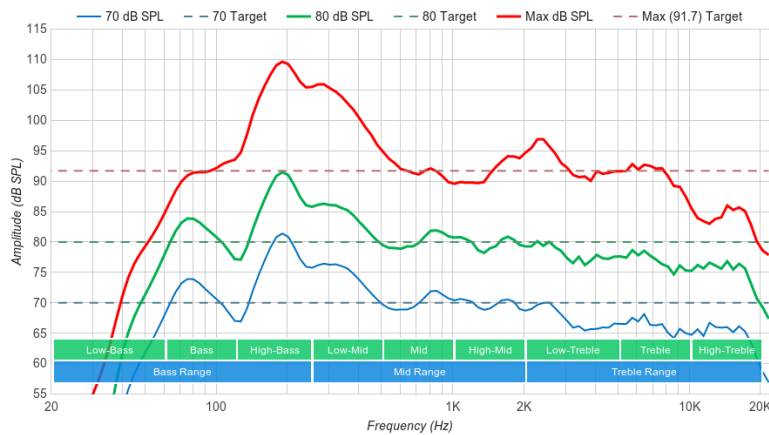


From "Master handbook of acoustics" by F. A. Everest & K. C. Pohlmann, 2009, McGraw-Hill Education.

By operating significantly above these, shown in Figures 3 and 4, the system avoids direct competition with the main noise sources, increasing the SNR. Being at the threshold of human hearing (Everest & Pohlmann, 2021), the transmission is discrete. However, validating consumer hardware (TV speakers and smartphone microphones) in this range is necessary. Market analysis shows that operational capacity is sufficient.

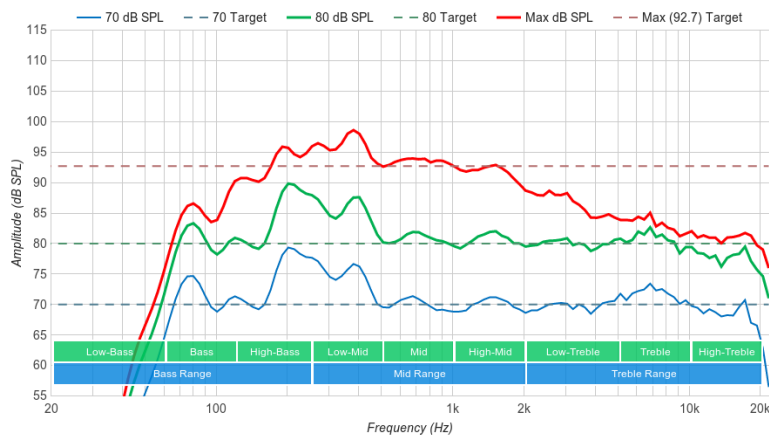
On the emitter side, consumer speakers like TVs have roll-off in high-treble frequencies (above 10–20 kHz) (Vafaei, 2018). Nevertheless, this performance proved functional. Figures 5 and 6 show frequency responses for Samsung and Sony models.

Figure 5 - Frequency response - SAMSUNG Q8C.



From "Frequency response of tvs" by S. Vafaei, 2018.

Figure 6 - Frequency response - Sony A1E OLE.

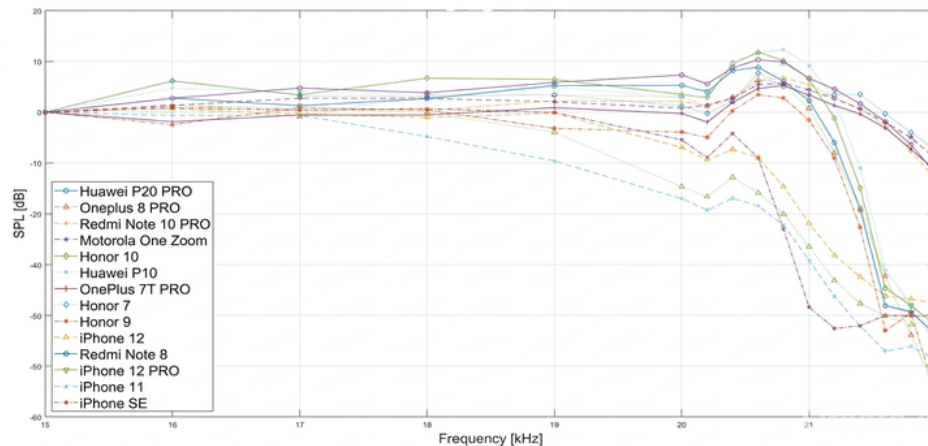


From "Frequency response of tvs" by S. Vafaei, 2018.

Hardware limitations influenced the modulation choice. State-of-the-art protocols like ggwave (Gerganov, 2022) use MFSK with up to 6 simultaneous frequencies. However, complex wave reproduction at these limits exceeds consumer speaker capacity. Tormim (2023) reported the unfeasibility of ultrasonic mode using internal PC speakers, reverting to the audible spectrum. In contrast, the serial FSK approach (one tone at a time) requires only simple spectral peak reproduction, bypassing intermodulation distortion.

Encouragingly, smartphone microphone tests, Figure 7, indicate that popular models maintain effective sensitivity up to 20 kHz. Hromadová et al. (2022) analyzed 14 smartphones, demonstrating that all captured signals reliably up to at least 20.6 kHz.

Figure 7 - Frequency response of microphones in 14 smartphone models.

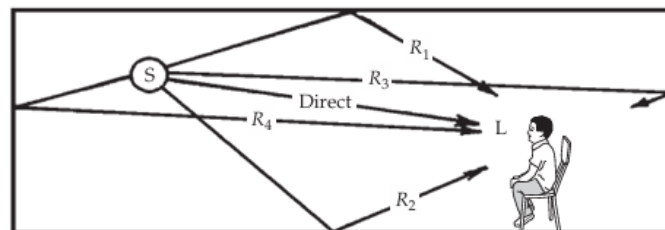


From “Frequency response of smartphones at the upper limit of the audible range” by V. Hromadová, P. Kasák, R. Jarina & P. Brida, 2022, ELEKTRO.

Treatment of reverberation and intersymbol interference

As reverberation is unavoidable in closed environments (Everest & Pohlmann, 2021), the source sound reaches the receiver via direct paths and multiple reflections, as illustrated in Figure 8.

Figure 8 - Direct sound arrives first, followed by reflected components (echo).



From “Master handbook of acoustics” by F. A. Everest & K. C. Pohlmann, 2021, McGraw-Hill Education.

The protocol’s main strategy was to mitigate ISI effects rather than eliminate them. Waiting for echoes to vanish would require silence intervals near the Reverberation Time (RT), making transmission too slow. Instead, the receiver only needs the current bit’s energy peak to be stronger than residual echoes. To achieve this, a Guard Interval was included: a silence period after each data symbol. This ensures temporal separation, allowing previous bit echoes to start decaying (Everest & Pohlmann, 2021).

Justification of processing and decoding architecture

Own architecture development was chosen over consolidated libraries like ggwave. While these represent the state of the art, they are designed for high data density via MFSK, assuming high receiver computational power. For OAuth architecture involving short tokens, computational efficiency and simplicity are priorities. The first divergence is frequency extraction. High-density protocols rely on the Fast Fourier Transform (FFT), which has a computational complexity of $O(N \log_2 N)$ (Bouchaleun, 2019). For a standard 1,024-sample buffer, this requires approximately 5,120 complex multiplications to calculate an entire spectrum, most of

which is irrelevant for this application. Since our Serial FSK protocol switches between only two frequencies, the FFT is inefficient (Puhmann, 2015). Instead, an optimized Goertzel Algorithm was implemented as a real-time resonant IIR filter. This reduces the complexity to $O(N)$ per frequency (Massachusetts Institute of Technology, 2006), requiring only 2,048 operations for the same buffer, a reduction of over 60% in processing load.

Libraries like `ggwave` use continuous transmission and heavy Reed-Solomon (ECC) error correction to overcome acoustic interference. As observed by Tormim (2023), these methods are computationally expensive and tailored for high-throughput streams.

Our system, designed for 32 or 64-bit tokens, prevents errors in the time domain: a Guard Interval (brief silence) waits for acoustic echo decay before the next bit is processed. This approach trades raw transmission speed for robustness and extreme power efficiency, as it avoids the CPU-intensive polynomial divisions required by Reed-Solomon decoders, which is critical for background authentication tasks on smartphones.

Results and discussion

Experimental scenario configuration

A controlled environment (5×3 meters) was used for validation. Hardware was divided into emitting and receiving stations. The emitter was a notebook (Intel Core i5, 16 GB RAM) running modulation code. Audio was sent via HDMI to a 32-inch Samsung LED TV, calibrated to 60 dB SPL. The receiver was a Samsung Galaxy A54 smartphone. Sound pressure measurements were taken at the receiver's position using a digital sound level meter.

Testing design

Tests studied signal integrity and success rate under different conditions (30 samples per configuration). Test vectors varied three parameters: **Emitter-Receiver Distance** (1.0 m, 2.3 m, and 3.0 m), **Payload Size** (32 bits and 64 bits), and **Transmission Rate** (20, 30, and 40 bps).

Two environmental conditions were tested:

1. **Condition A (Low Noise):** Closed doors/windows, no artificial sources (approx. 30 dB SPL).
2. **Condition B (High Noise/Interference):** Open doors/windows, with an external speaker (AMVOX 220W) playing music (75–80 dB SPL).

Frequency responses were measured via the Spectroid app. Figure 9 shows Condition A. Figure 10 shows an 18 kHz peak during operation.

Figure 9 - Ambient frequency response graph in Condition A with system off.

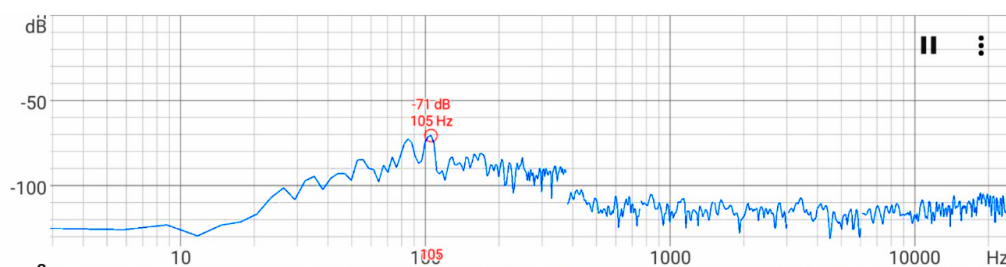


Figure 10 - Ambient frequency response graph in Condition A with system operating.

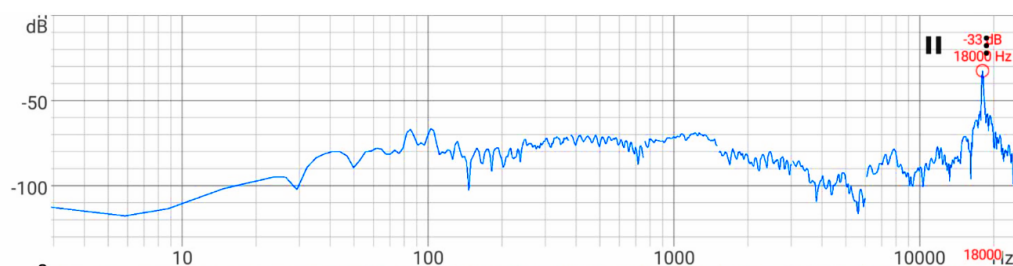


Figure 11 shows high noise in Condition B (100 Hz–10 kHz). Figure 12 shows the 18 kHz peak overlapping with the noise.

Figure 11 - Ambient frequency response graph in Condition B with system off.

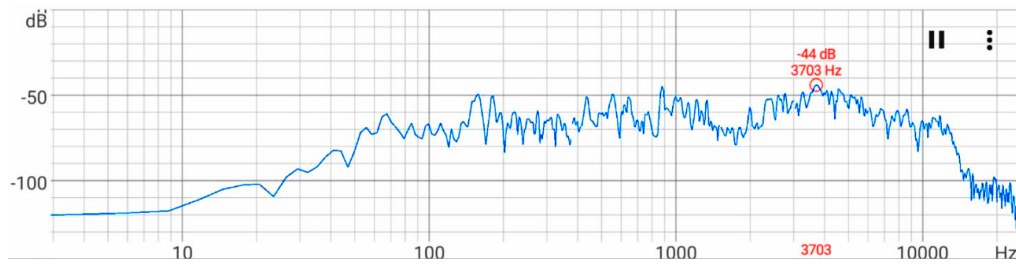
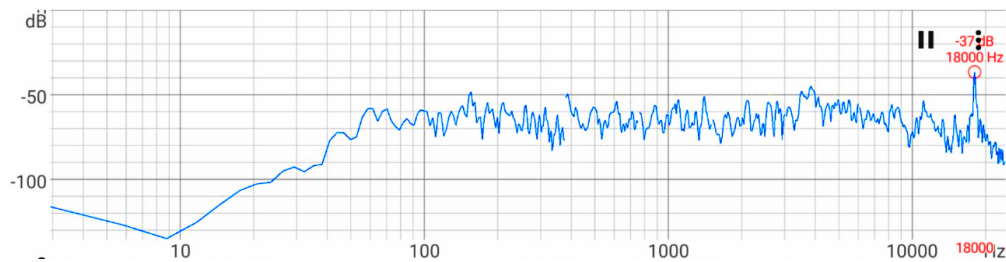


Figure 12 - Ambient frequency response graph in Condition B with system operating.



Tables 2 to 3 show the test results.

Table 2 - Transmission performance under Condition A.

Distance (m)	Packet (bits)	Bitrate (bps)	Success Rate (%)	Time (s)
1.0	32	20	83.3	5.450
	32	30	76.6	3.800
	32	40	53.3	2.975
	64	20	73.3	10.250
	64	30	53.3	7.000
	64	40	33.3	5.375
2.3	32	20	96.6	5.450
	32	30	86.6	3.800
	32	40	63.3	2.975
	64	20	83.3	10.250
	64	30	76.6	7.000
	64	40	53.3	5.375
3.0	32	20	96.6	5.450
	32	30	73.3	3.800
	32	40	60.0	2.975
	64	20	80.0	10.250
	64	30	70.0	7.000
	64	40	53.0	5.375

Table 3 - Transmission performance under Condition B.

Distance (m)	Packet (bits)	Bitrate (bps)	Success Rate (%)	Time (s)
1.0	32	20	83.3	5.450
	32	30	66.6	3.800
	32	40	60.0	2.975
	64	20	80.0	10.250
	64	30	56.6	7.000
	64	40	30.0	5.375
2.3	32	20	96.6	5.450
	32	30	83.3	3.800
	32	40	56.6	2.975
	64	20	80.0	10.250
	64	30	66.6	7.000
	64	40	60.0	5.375
3.0	32	20	96.6	5.450
	32	30	80.0	3.800
	32	40	73.3	2.975
	64	20	90.0	10.250
	64	30	70.0	7.000
	64	40	50.0	5.375

Statistical analysis of results

A Generalized Linear Model (GLM) with binomial distribution and logit link function was used (Paula, 2025). Distance, Packet, Bitrate, and Condition were treated as categorical factors. The relationship between the explanatory variables and the probability of success is expressed by:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \sum_j \beta_j X_{ij}, \quad (8)$$

where the coefficients β_j quantify the effect of the explanatory variables on the log-odds of success.

A successful transmission was explicitly defined as the correct reception and full decoding of the transmitted token, with no bit errors. Thus, for each experimental configuration:

$$Y_i \sim \text{Binomial}(n = 30, p_i),$$

where p_i denotes the probability of success for the i -th configuration.

The variables *Distance*, *Packet*, *Bitrate*, and *Condition* were treated as categorical factors, with effects estimated relative to reference categories. Table 4 presents the estimated odds ratios and their corresponding 95% confidence intervals.

Table 4 - Estimated odds ratios and statistical significance of system variables.

Variable	Odds Ratio	95% CI		p-value
		Lower	Upper	
Intercept	6.2821	4.1655	9.4743	< 0.001
Distance (2.3 m)	1.9743	1.4032	2.7778	< 0.001
Distance (3.0 m)	1.8793	1.3383	2.6390	< 0.001
Packet (64 bits)	0.5017	0.3778	0.6661	< 0.001
Bitrate (30 bps)	0.3763	0.2555	0.5544	< 0.001
Bitrate (40 bps)	0.1673	0.1148	0.2439	< 0.001
Condition (2)	1.0418	0.7871	1.3789	0.775

The results indicate that *Distance*, *Packet size*, and *Bitrate* have statistically significant effects on the response variable (p -values < 0.05). Specifically, the variable *Condition* was not statistically significant ($p > 0.05$), suggesting no measurable effect of ambient noise under the tested conditions.

Analysis indicated that increasing packet size and, primarily, bitrate are the most significant factors in success rate degradation. Counter-intuitively, performance at short distances (1.0 m) was lower than at moderate distances (2.3 m and 3.0 m), likely due to interference where previous signals were not sufficiently mitigated. Noise variation did not statistically influence performance, proving robustness.

Conversely, intermediate distances (2.3 m and 3.0 m) yielded better performance compared to the shortest distance (1.0 m), suggesting the presence of interference effects at very short ranges, possibly due to insufficient signal dissipation between consecutive transmissions.

The lack of statistical significance for the *Condition* variable suggests robustness to ambient noise under the evaluated scenarios.

Finally, regarding the use of more complex modulation schemes such as MFSK, their practical adoption may be limited in this context due to spectral constraints, increased inter-frequency interference, and higher computational requirements for detection. Therefore, simpler modulation schemes represent a more suitable trade-off between performance and implementation feasibility.

Conclusions

This study successfully demonstrated the development and validation of an alternative, low-overhead acoustic authentication protocol designed to seamlessly bridge the login credential transmission gap between primary limited-input emitters and secondary receiving devices. By replacing friction-heavy manual typing or visual alignment dependencies with non-audible sound tokens, the system presents a highly functional, discrete, and potentially accessible solution that directly enhances hardware usability for consumer technology.

Author Contributions

B. H. F. Stranak: Conceptualization, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **J. C. Pizolato Junior:** Conceptualization, Project administration, Supervision. **F. A. Kitazawa:** Formal analysis.

Conflicts of Interest

The authors declare no conflict of interest.

References

- Bistafa, S. R. (2018). *Acústica Aplicada ao Controle do Ruído* (3rd ed.). Blucher.
- Bouchaleun, A. (2019). An Elementary Introduction to Fast Fourier Transform Algorithms. <https://math.uchicago.edu/~may/REU2019/REUPapers/Bouchaleun.pdf>
- Denniss, W., Bradley, J., Jones, M. B., & Tschofenig, H. (2019). *OAuth 2.0 Device Authorization Grant* (RFC No. 8628). Internet Engineering Task Force. <https://doi.org/10.17487/RFC8628>
- Everest, F. A., & Pohlmann, K. C. (2021). *Master Handbook of Acoustics* (7th ed.). McGraw-Hill Education.
- Gerganov, G. (2022). *Ggwave: Ggwave github projectI* [Computer code]. GitHub. <https://github.com/ggerganov/ggwave/blob/master/include/ggwave/ggwave.h>
- Google Developers. (2025). *OAuth 2.0 for tv and limited-input device applications* [Application Programming Interface]. <https://developers.google.com/identity/protocols/oauth2/limited-input-device?hl=pt-br>
- Guimarães, D. A., & de Souza, R. A. A. (2015). Exploring Practical Aspects of the FSK Modulation with Non-Coherent Matched Filter Detection. *Revista de Tecnologia da Informação e Comunicação*, 5(1), 17–23. <https://inatel.br/docentes/documents/dayan/Publications/77.pdf>

- Hromadová, V., Kasák, P., Jarina, R., & Brida, P. (2022). Frequency Response of Smartphones at the Upper Limit of the Audible Range. In Institute of Electrical and Electronics Engineers, *Conferences [Proceedings]*. 2022 ELEKTRO, Krakow, Poland. <https://ieeexplore.ieee.org/document/9803475>
- Massachusetts Institute of Technology. (2006). *Lecture 20: The Goertzel Algorithm and the Chirp Transform*. MIT OpenCourseWare, 6.341 Discrete-Time Signal Processing. <https://ocw.mit.edu/courses/6-341-discrete-time-signal-processing-fall-2005/resources/lec20/>
- Nielsen, J. (1993). *Usability Engineering*. Academic Press. <https://dl.acm.org/doi/pdf/10.5555/2821575>
- Paula, G. A. (2025). *Modelos de Regressão*. Instituto de Matemática e Estatística, Universidade de São Paulo. <http://www.ime.usp.br/~giapaula/textoregressao.htm>
- Proakis, J. G., & Salehi, M. (2008). *Digital Communications* (5th ed.). McGraw-Hill.
- Puhlmann, H. F. W. (2015). *Transformada Discreta de Fourier – Algoritmo de Goertzel*. ResearchGate. <https://www.researchgate.net/publication/281034964>
- Tormim, M. R. (2023). *Protocolo de Transmissão por Proximidade com Dados sobre Som* [Trabalho de Conclusão de Curso, Universidade Federal de Catalão]. https://prod.ufcat.edu.br:1337/uploads/monografia_pfc_FINAL_marcelo_ribeiro_tormim_8193942231.pdf
- Vafaei, S. (2018). Frequency Response of TVs. <https://www.rtings.com/tv/tests/sound-quality/frequency-response>
- Vieira, G. F., & Costa, E. d. S. (2012). Avaliação Acústica de Ambientes Fechados: Estudos e Medições. medições. In Associação Brasileira de Engenharia e Ciências Mecânicas, *CONEM 2012 Engenharia em Destaque* [Anais]. 7º Congresso Nacional de Engenharia Mecânica, São Luís, Maranhão, Brasil. <https://www.abcm.org.br/anais/conem/2012/PDF/CONEM2012-1839.pdf>
- Watson, B. (2001). *FSK: Signals and Demodulation*. (Tech-note). WJ Communications. (Originally published in Watkins-Johnson Company Tech-notes Vol. 7 n. 5, 1980). https://pt.scribd.com/document/781910367/0-WJ-Tech-Note-Vol-7-No-5#google_vignette&content=query:2001,pageNum:8,indexOnPage:0,bestMatch:false