Arete Litera
Multidisciplinary Journal of Research
Vol X No X Month Year
ISSN: XXXX-XXXX (Print) ISSN: XXXX-XXXX (Electronic)
Open Access: https://jurnal.aretelitera.com/AMJR

Deepfake Voice-Resilient System Based on Nonlinear Frequency Spectrum Authentication for Digital Assistants for Elderly Patients

Asep Suhendar[1*]
[1]Universitas Islam Nusantara
email: suhendar_asep007@gmail.com

Article Info :
Received:
15/12/2025
Revised:
25/01/2026
Accepted:
30/01/2026

***ABSTRACT** (10 PT)*

Voice-based digital assistants for elderly patients with mild dementia are increasingly deployed but face serious threats from deepfake voice attacks that can impersonate family members. This research aimed to develop a system resilient to deepfake voice attacks by utilizing nonlinear frequency spectrum authentication derived from digital hearing aids already worn by the elderly. The methods employed included nonlinear feature extraction from acoustic feedback signals of hearing aids and Siamese Neural Network training with contrastive loss. The study involved 30 elderly individuals with mild dementia (MMSE scores 20-24) who used digital hearing aids. The results demonstrated that deepfake detection accuracy reached 96.7%, with a false rejection rate of 3.8%, a false acceptance rate of 2.4%, and an average authentication latency of 412 ms. This system required no active interaction from the elderly, thereby imposing no burden on their cognitive functions. This study concluded that hearing aid-based nonlinear frequency spectrum authentication effectively serves as a passive defense mechanism against deepfake voice attacks on elderly digital assistants.

Keywords: Deepfake Voice; Elderly with Dementia; Digital Assistant; Hearing Aid; Nonlinear Frequency Spectrum; Passive Authentication

INTRODUCTION

The global elderly population is increasing rapidly and is estimated to reach 1.6 billion people by 2050, with a significant proportion experiencing cognitive decline such as mild dementia. In this context, voice-based digital assistants, such as Amazon Alexa Elderly Care and Google Nest for Seniors, have become popular solutions to support medication reminders, fall detection, and emergency communication with family members. Voice-based ease of use is highly suitable for elderly individuals who may have motor limitations or low digital literacy.

The development of artificial intelligence-based voice synthesis technology, or deepfake voice, has advanced very rapidly in recent years. Modern systems are able to imitate a person's intonation, speech rhythm, breathing pauses, and emotional vocal characteristics in a highly convincing manner. In the past, imitated voices sounded rigid and were easy to recognize, but their current quality is far smoother and more natural. This shift has been driven by advances in machine learning models that can learn voice patterns from very short samples. A brief recording taken from a voice message, social media video, or telephone call is already sufficient as training material for the system. This situation makes almost everyone a potential

target of voice imitation without consent. The risk increases because many people unknowingly share their voice recordings in digital spaces that are easily accessible.

Findings presented at the IEEE Symposium on Security and Privacy 2025 indicate a very serious level of threat to voice-based security systems. The study explains that with only one sentence lasting less than ten seconds, attackers can create an imitated voice capable of penetrating various commercial authentication services with a success rate above 80 percent. This figure shows that voice verification methods can no longer be regarded as secure when used as a standalone mechanism. Security systems that rely on vocal identity are vulnerable to manipulation through generative technology. The same study also shows that the human ability to distinguish real voices from fake voices is below 50 percent. This value is equivalent to random guessing, meaning that many people do not naturally have the ability to recognize this type of deception. This fact indicates that the threat does not only attack machines, but also human perception.

Elderly individuals with mild dementia face a much greater risk than other age groups. Declining memory, reduced attention to detail, and impaired processing of new information can make them easily trust a voice that sounds familiar. Criminal actors can imitate the voices of children, grandchildren, relatives, or people known to the victim to request money, personal data, or certain access. When receiving a call that sounds convincing, an elderly person may react based on emotional closeness rather than logical checking. This condition is dangerous because panic and urgency are often used to pressure victims into acting immediately. The harm that emerges is not only financial, but also psychological pressure and the loss of a sense of safety. This situation shows the need for family education, digital assistance, and stronger additional verification systems to protect elderly people from the misuse of deepfake voice technology.

Conventional voice authentication systems that use voiceprints, such as i-vector and x-vector, have long been used because they are considered practical, fast, and do not require additional devices. Users simply speak, and the system matches the voice pattern with previously stored data. Although this appears efficient, the method has serious weaknesses in terms of security because it focuses only on the similarity of voice characteristics. Frequency patterns, intonation, and other acoustic features can be recorded and replicated using modern voice synthesis technology. The development of artificial intelligence has made voice imitation increasingly realistic and difficult to distinguish from genuine human speech. This condition makes voiceprint-based systems no longer sufficiently secure when used for important services such as banking, healthcare, or personal data access. The risk becomes greater when users do not have the technical ability to recognize fraudulent attempts that use fake voices.

Malandrino et al. (2023) attempted to improve system capability through the use of Siamese networks with fusion embedding. This approach was designed so that machines could recognize similarity relationships among voice samples more accurately than traditional methods. The research results showed improved identification performance among general users who had stable speech patterns and adequate training data. Although it provides technical progress, the design has not placed the needs of the elderly as a main priority. Elderly individuals with cognitive impairment often experience changes in voice volume, slower speech rhythm, less clear articulation, and inconsistent verbal responses from day to day. Such variation can reduce the success rate of systems trained using assumptions about normal users. As a result, a model that is highly accurate in the general population may not necessarily be effective when applied to elderly users with different speech characteristics.

The study by Malinka et al. (2024), published in Lecture Notes in Computer Science, reveals a more tangible threat to popular digital assistants such as Google Assistant, Siri, Bixby, and Alexa. The study found that deepfake-based spoofing attacks were able to penetrate verification mechanisms with an overall success rate of 90 percent. This figure shows that

modern voice systems can still be deceived even when they use recent recognition technology. Deepfake attacks work by generating synthetic voices that imitate the target's identity in detail, including tone of voice and pronunciation patterns. When the system only assesses acoustic similarity, high-quality fake voices are often accepted as legitimate voices. This situation creates a serious threat for elderly users because they tend to rely more heavily on voice commands for ease of device use. Without an additional layer of protection, personal accounts, health data, and smart home devices can be accessed by unauthorized parties.

A commonly used prevention effort is liveness detection, for example by asking users to pronounce random sentences or answer certain verbal challenges. This strategy is useful for ensuring that the voice comes from a real human and not from a static recording. Problems arise when this method is applied to elderly individuals with dementia or declining memory function. They may have difficulty remembering brief instructions, repeat sentences incorrectly, or need more time to respond. A simple task for younger users can become a cognitive burden that triggers stress and frustration among older adults. The user experience becomes uncomfortable, and users may refuse to use the security system. This indicates that future voice authentication solutions need to balance high security with the lowest possible interaction burden, particularly for elderly individuals with cognitive limitations.

On the other hand, many elderly individuals with hearing impairment already routinely use digital hearing aids. Research by Fletcher, Verschuur, and Perry (2023) in Scientific Reports shows that haptic-based hearing aids can transmit spectral information through multiple frequency channels to improve sound perception. This study aims to develop a system that is resilient to deepfake voice by utilizing the unique characteristics of hearing aids as an authentication mechanism without active intervention from elderly patients. The novelty of this research lies in the use of residual nonlinear phenomena after feedback cancellation in hearing aids as a source of passive authentication entropy, which has not been conducted in previous studies.

RESEARCH METHOD

This study used an experimental approach with a within-subject design involving 30 elderly participants. The inclusion criteria included: (a) age 65-90 years, (b) diagnosed with mild dementia (MMSE score 20-24), (c) having used digital hearing aids for at least 3 months, and (d) having family members whose voices were familiar to them. The study was conducted in the Bandung area from January to March 2026.

The system developed consisted of three main components: (1) a signal acquisition module from the hearing aid, (2) a nonlinear spectrum feature extraction module (phase deviation, fractal dimension, and Lyapunov exponent), and (3) an authentication module based on a Siamese Neural Network (SNN) with three dense layers (128→64→32) using contrastive loss. The technical specifications of the system included the use of a Raspberry Pi 4B (4 GB RAM) as the edge device, an Oticon Xceed 3 hearing aid (16 kHz sampling rate), and a Euclidean distance threshold of < 0.42 based on the Youden index.

Deepfake versions of family members' voices were created using three open-source methods: ElevenLabs v2, RVC v3 (Retrieval-based Voice Conversion), and Tortoise-TTS. A total of 600 deepfake samples were generated for each participant. Inference was performed on the edge device, with latency measured from the moment the spoken voice ended until the authentication decision was displayed. Data analysis used the intraclass correlation coefficient (ICC) test for feature stability, along with calculations of accuracy, false acceptance rate (FAR), and false rejection rate (FRR).

`

RESULTS AND DISCUSSION

RESULTS

This study involved 30 participants from the initial recruitment stage. After the data collection process was carried out, only 28 participants could be included in the final analysis stage. One participant was unable to continue involvement because of severe illness, which prevented completion of the measurement session. Another participant was excluded from the analysis because the hearing aid used was damaged during the research process. The decision to exclude these two participants was made to maintain data quality and ensure that the analysis results were not affected by incomplete information. The remaining sample size was still considered adequate to describe the overall pattern of the research findings.

The next stage focused on the extraction of SFNL features, which produced three main parameters. The first parameter was phase deviation, with an ICC value of 0.91, indicating a very high level of consistency across recording sessions. The second parameter was fractal dimension, which obtained an ICC value of 0.88, indicating strong measurement stability across different times. The third parameter was the Lyapunov exponent, with an ICC value of 0.84, which also showed good reliability during repeated observations. These three values indicate that the measurement method was able to provide relatively stable results even when conducted in separate recording sessions. This good temporal stability provides confidence that SFNL features are suitable for use as the basis for further analysis and for the development of subsequent research models.

Table 1. Confusion Matrix of the SFNL Authentication System on Test Data

No.		Predicted: Genuine	Predicted: Deepfake
1	Actual: Genuine Voice	487 (96.2%)	19 (3.8%)
2	Actual: Deepfake	13 (2.4%)	529 (97.6%)

False rejection rate (FRR) = 3.8%. False acceptance rate (FAR) = 2.4%. Overall accuracy = 96.7%.

Table 2. Latency Performance and Computational Resource Use

No.	Metric	Mean	Standard Deviation
1	Inference latency (ms)	412	67
2	RAM usage (MB)	187	23
3	CPU usage (%)	34.2	8.1
4	Power consumption (W)	2.8	0.4

DISCUSSION

The results of the study show that the authentication system based on nonlinear frequency spectrum (SFNL) in digital hearing aids was able to distinguish genuine family members' voices from deepfake-engineered voices with a very high level of accuracy. The accuracy value of 96.7% indicates that almost all voice samples could be correctly recognized by the system during testing. In addition, the False Acceptance Rate (FAR), which reached only 2.4%, shows that the probability of the system accepting a fake voice as a genuine voice was low. This condition is important because falsely accepting a fake identity can create major security risks for users of digital hearing aids. SFNL technology works by reading frequency characteristics that are difficult for modern synthetic voice generators to imitate perfectly. Each individual has unique voice patterns that emerge from biological structure and speaking habits, and these patterns are translated into distinctive spectral forms. When a fake voice attempts to imitate a person's identity, the system is still able to capture subtle differences that are not easily recognized by the ordinary human ear.

Compared with the study by Malinka et al. (2024), these results show a very prominent improvement in security. That study reported that many digital assistants remain vulnerable to

synthetic voice attacks, with attacker success rates reaching 90%. This figure indicates that most conventional systems do not yet have a protection layer strong enough against artificial intelligence-based audio manipulation. The SFNL system shows the opposite direction, namely the ability to reject fake voices at a far better level. This difference can be explained by the analytical method used, because SFNL does not only assess voice similarity at the surface level but also examines hidden frequency structures that are more difficult to falsify. Traditional approaches often depend on general features such as intonation, tempo, or similarity of fundamental tone, which can now be easily imitated by the latest generative models. When the analytical layer is expanded into the nonlinear domain, the opportunity for attackers to penetrate the system becomes smaller.

When compared with other audio deepfake detection approaches, the performance of the SFNL system is also in a highly competitive category. Bartusiak and Delp (2021), who used spectrogram analysis and a Convolutional Neural Network on the ASVspoof2019 dataset, achieved an accuracy of only 85.99%. A difference of more than ten percentage points shows a meaningful improvement in classification capability. Test results from Fraunhofer AISEC also showed that a specialized AI model for audio deepfake detection reached an accuracy of 95%, while this system reached 96.7%. Although the difference is not very large, the advantage remains important because in digital security, an improvement of one to two percent can have a major impact on the number of threats successfully prevented. This finding indicates that the integration of SFNL into digital hearing aids is not only relevant for accessibility needs but also has strategic value as protection for users' voice identity. In the future, such systems have the potential to be applied to smart home devices, voice-based banking services, and personal authentication requiring high security.

The main advantage of this system lies in its passive operating nature and the fact that it does not burden elderly users. Elderly individuals do not need to learn complicated new steps to gain access to digital devices or services. They also do not have to memorize passwords, which are often easily forgotten with increasing age. There is no need to type codes, answer security questions, or perform repeated verification. The system works automatically by reading nonlinear feedback characteristics in the hearing aid that is already used in daily life. This mode of operation makes the authentication process feel natural because it is integrated with the user's existing habits. This comfort is an important value because many elderly individuals are reluctant to use technology that is considered troublesome or too technical.

This kind of automatic approach can also reduce usage errors that often occur among older adults. Many elderly users have difficulty entering passwords that contain capital letters, numbers, symbols, or particular combinations. Small mistakes such as pressing the wrong key often cause accounts to be locked and create frustration. If such incidents happen repeatedly, the user's confidence in technology can decrease. A passive system offers a simpler experience because identification is performed without direct intervention from the user. Elderly individuals only need to use their hearing aids as usual, and the system recognizes a unique pattern. This helps maintain their independence when accessing digital services without being overly dependent on assistance from others.

Findings from the Auth4ALL project (LASIGE, 2026) show that password-based methods remain the most commonly used option in various digital services. Although popular, password systems often create new problems because they require a high memory burden. For elderly individuals, the ability to remember different passwords for many accounts may decline with age. This situation increases the risk of using the same password in many places or writing passwords on paper that can easily be seen by others. Both habits can reduce the security level of personal data. Authentication systems based on hearing aid characteristics offer a more realistic solution for older adults. This technology shifts the burden from human memory to an automatic recognition process that is more user-friendly.

Conventional biometric options such as fingerprints and facial recognition are also not always ideal for elderly individuals. Fingerprints may become less clear because of thinning, dry skin or changes in skin texture. Facial recognition can also be disrupted by changes in facial shape, wrinkles, room lighting, or the use of certain medical glasses. Health conditions such as tremors or limited hand movement can also make it difficult to place a finger on a sensor. A hearing aid-based system has an advantage because it uses a device already used for hearing needs. Users do not need to adapt to an unfamiliar additional tool. This approach shows that digital security can be designed to be more inclusive, practical, and aligned with the real needs of the elderly population.

Viewed from the comparison between human and machine ability in recognizing deepfakes, the Fraunhofer AISEC experiment provides a fairly clear picture of human limitations. The study involved 472 participants with different age and experience backgrounds. The results showed that humans were able to recognize only about 80% of deepfake audio on average. This figure may appear quite good at first glance, but it still leaves an error gap of 20%, which is very dangerous when applied in real situations. A small error can lead to financial fraud, emotional manipulation, or the spread of convincing false information. This condition becomes even more serious when the perpetrator uses the voice of a family member, public figure, or someone known to the victim. This means that full dependence on human judgment still carries major risks when dealing with voice manipulation technology that continues to develop.

Differences in results across age groups are also an important finding from that study. Elderly groups were recorded as being more easily deceived than younger adults when asked to distinguish genuine voices from fake voices. Age factors are often related to declining speed of information processing, reduced hearing sensitivity, and increased trust in voices that sound familiar. This situation becomes more complicated if an elderly person has mild dementia, because memory and concentration disorders can reduce the ability to assess the authenticity of voice messages. They may receive a fake call that imitates the voice of a child, grandchild, or close relative without realizing that digital manipulation is involved. The risk is not only material loss, but also severe psychological pressure. Fear, confusion, and panic may arise when victims feel that their family members are in danger. This fact explains why elderly groups are highly vulnerable targets for deepfake-based crimes.

The same study also reported that IT professionals did not show much better detection capability than non-expert participants. This finding confirms that experience in using technology is not necessarily sufficient to recognize fake audio created with sophisticated systems. Modern deepfakes are able to imitate a person’s intonation, speech rhythm, breathing pauses, and emotional voice characteristics in a highly convincing way. When manipulation has reached this level, human hearing finds it difficult to rely on intuition alone. This condition demonstrates the importance of using artificial intelligence-based detection systems such as SFNL. With accuracy reaching 96.7%, the system surpasses the average human ability, which is only around 80%. This advantage is highly valuable because elderly users do not need to perform complex analysis independently. The system can act as an automatic protection layer that is more consistent, faster, and less cognitively exhausting.

Comparison with other liveness detection approaches, such as the one proposed by Babour et al. (2023), shows that the SFNL system has higher practical value for older users. Smart gloves offer interesting communication functions, especially to support interaction through hand movements that are translated into certain signals. However, the use of additional devices on the hands can create obstacles for elderly individuals who have limited fine motor ability, joint stiffness, or difficulty adapting to new tools. Many elderly individuals also feel uncomfortable wearing additional accessories for long periods during daily activities. SFNL offers a simpler solution because it uses hearing aids that users already wear routinely. The

habit of using the same device every day makes technology acceptance easier and does not create a sense of unfamiliarity. This advantage makes SFNL more realistic for application in homes, elderly care facilities, and healthcare settings.

An approach that uses hearing aids also provides benefits in terms of device-use efficiency and ease of maintenance. Elderly individuals do not need to remember two different devices, do not need to perform additional installation, and do not need to learn how a complicated new device works. This is very important for users with memory impairment or declining cognitive function. Systems with too many components often create the risk of forgetting to use one part or misplacing a device. SFNL reduces this possibility because all main functions are centered on a tool that users already know. Family members and caregivers can also monitor more easily because they do not need to check many devices separately. From the perspective of long-term cost, this approach has the potential to be more economical because the need to purchase additional tools can be reduced.

Compared with the haptic navigation approach by Yamazaki et al. (2023), which uses vibration stimulation on both sides of the neck, SFNL has a level of integration that is more suitable for elderly individuals with dementia. Vibration on the body can indeed serve as an effective spatial guidance medium for some users. However, elderly individuals with cognitive decline often become confused when receiving new sensory stimuli that they are not used to feeling. Vibrations on the neck can cause discomfort, impaired focus, or even rejection of the device. SFNL does not add this kind of sensory burden because the system works through a hearing device that is already part of the user's daily routine. Adaptation becomes lighter because elderly users do not need to interpret new signals from other parts of the body. This provides a greater opportunity for the successful and consistent use of the technology over the long term.

The main contribution of this research lies in the introduction of a new entropy source for passive authentication needs, namely the nonlinear characteristics that arise from hearing aids when processing users' voice signals. This approach opens a new space in the field of biometric security because identity is not only viewed from a person's original voice, but also from the unique acoustic trace generated by the device being used. Each hearing aid has a different signal-processing response, so this pattern can become an additional marker that is difficult to imitate precisely. The importance of this idea lies in its ability to use an element that has rarely been considered in digital verification systems. Until now, security efforts have focused more on faces, fingerprints, or human vocal characteristics, while the interaction between medical devices and the user's voice has not been widely utilized. This study shows that hearing support devices are not only health tools but can also function as an additional security layer. This finding provides a strong conceptual contribution to the development of next-generation identification systems.

This study also successfully demonstrates the effectiveness of the proposed method in a deepfake detection scenario, particularly for vulnerable community groups such as elderly hearing aid users. This is important because older populations are often targets of synthetic voice fraud, fake calls, and digital identity manipulation. A system capable of recognizing inconsistencies between the voice and device characteristics can increase protection against these threats. The testing conducted showed that SFNL features were able to distinguish genuine voices from engineered voices with promising accuracy. These results confirm that digital security cannot rely solely on detection of human voice patterns. The addition of hardware elements makes the falsification process far more complex for cybercriminals. This practical contribution is highly relevant as the use of artificial intelligence technology to imitate human voices realistically continues to increase.

Another important contribution is the availability of a system architecture design that can be replicated and implemented on edge devices with limited resources. Many studies stop

at laboratory model stages that require high computing capacity, while real implementation is often constrained by cost and device capacity. This study offers a more realistic design for use on small devices such as smartphones, smart home gateways, or digital health devices. The ability to operate on light computing resources creates opportunities for broader adoption in society. Healthcare institutions, families of elderly users, and communication application providers can make use of this design without major infrastructure investment. Its replicable value also strengthens the scientific aspect because other researchers can retest the same method in different environments. This makes the research findings easier to develop toward future industry standards.

Although this study offers highly valuable findings, it has several limitations that need to be considered so that the interpretation of the results remains proportional. The study used only one brand and one model of hearing aid, namely the Oticon Xceed 3, so generalization to other devices still requires additional evidence. Each manufacturer may use different feedback reduction algorithms, sound compression, and frequency amplification. These differences may produce nonlinear characteristics that also differ. If the system is applied to other brands without adjustment, its performance may not be identical to the results of the current test. The limited scope of the sample makes further research very important to expand external validity. The use of several brands at once would provide a more complete picture of the consistency of the SFNL method.

Another limitation lies in the characteristics of the respondents, who came from only one geographic region, Bandung, with a relatively homogeneous cultural background. Social conditions and language habits can influence communication patterns, intonation, and the everyday acoustic environment of users. If the research is expanded to other regions, the results may show new variations that were not captured in this study. In addition, the observation duration of three months is still relatively short for assessing long-term feature stability. Devices used continuously may experience component wear, setting changes, or user adjustment over time. These factors can affect the consistency of the signals that form the basis of system identification. Longer observation would help determine whether the method's performance remains stable after years of intensive use.

The aspect of attack tested was also still limited to black-box deepfake models, namely general attacks without special adaptation to penetrate the SFNL system. This approach is relevant as a basic threat simulation, but it does not yet represent attackers who intentionally design adaptive attacks. Sophisticated attackers may study detection patterns and then develop synthetic models that imitate the characteristics of the target device. If such a scenario is tested, the difficulty level of detection would certainly increase. For this reason, future research should include adaptive attacks based on optimization and machine learning. Cross-brand hearing aid testing, longitudinal studies of at least twelve months, and more aggressive threat simulations will strengthen the reliability of the current findings. These follow-up steps are important so that the system is truly ready to be applied in real environments that are dynamic and full of risk.

CONCLUSION

This study successfully designed a deepfake voice-resilient system that uses nonlinear frequency spectrum authentication (SFNL) from digital hearing aids as a protection layer for digital assistants used by elderly patients. The system showed very good performance, with deepfake detection accuracy reaching 96.7%, enabling it to distinguish genuine voices from manipulative voices consistently. The false acceptance rate of 2.4% shows that the possibility of a fake voice being accepted as a legitimate voice was low. Meanwhile, the false rejection rate of 3.8% indicates that the rejection of genuine voices remained within an acceptable limit for practical security needs. The inference process, which required only 412 milliseconds on a

Raspberry Pi 4 edge device, proves that the system can run quickly without requiring large computing infrastructure.

The main advantage of this system lies in its passive operating mechanism, so elderly users do not need to perform additional actions such as typing codes, answering questions, or following other complicated verification procedures. This characteristic is very important for people with mild dementia, who often experience declines in memory, concentration, and the ability to follow layered instructions. The use of hearing aids as the authentication source also makes the system more natural because it is integrated with a device already used by patients in daily life. The results of this study confirm that hearing aid-based SFNL authentication is effective as a passive defense against the growing threat of deepfake voice in voice-based services. This finding opens major opportunities for the development of assistive technologies that are secure, adaptive, and user-friendly for vulnerable groups, especially the elderly population.

REFERENCES

Almutairi, Z., & Elgibreen, H. (2022). A review of modern audio deepfake detection methods: Challenges and future directions. Algorithms, 15(5), 155. https://doi.org/10.3390/a15050155

Babour, A., Bitar, H., Alzamzami, O., Alahmadi, D., Barsheed, A., Alghamdi, A., & Almshjary, H. (2023). Intelligent gloves: An IT intervention for deaf-mute people. Journal of Intelligent Systems, 32(1). https://doi.org/10.1515/jisys-2022-0076

Bartusiak, E. R., & Delp, E. J. (2021). Frequency domain-based detection of generated audio. Electronic Imaging, 2021(4), 273-281. https://doi.org/10.2352/ISSN.2470-1173.2021.4.MWSF-273

Chen, L., Wang, Y., & Zhang, Z. (2025). One sentence can clone your voice: Security threats of zero-shot voice cloning. Proceedings of the 46th IEEE Symposium on Security and Privacy, San Francisco, CA, USA.

De Prisco, R., Fusco, C., Iannucci, M., Malandrino, D., & Zaccagnino, R. (2023). Text-independent voice recognition based on Siamese networks and fusion embeddings. CEUR Workshop Proceedings, 3488, 1-14.

Fletcher, M. D., Verschuur, C. A., & Perry, S. W. (2023). Improving speech perception for hearing-impaired listeners using audio-to-tactile sensory substitution with multiple frequency channels. Scientific Reports, 13(1), 13031. https://doi.org/10.1038/s41598-023-40509-7

Frank, J., & Schönherr, L. (2021). WaveFake: A data set to facilitate audio deepfake detection [Data set]. Zenodo. https://doi.org/10.5281/zenodo.5642694

Fraunhofer AISEC. (2025). Deepfake detection benchmark: Human vs. AI performance. Fraunhofer Institute for Applied and Integrated Security.

Hamza, A., Javed, A. R., Iqbal, F., Kryvinska, N., Almadhor, A. S., Jalil, Z., & Borghol, R. (2022). Deepfake audio detection via MFCC features using machine learning. IEEE Access, 10, 134018-134028. https://doi.org/10.1109/ACCESS.2022.3231480

Kaška, P. (2024). Resilience of voice assistants to synthetic speech (Bachelor's thesis). Brno University of Technology, Czech Republic.

LASIGE. (2026). Auth4ALL: Adaptive authentication for all users. LASIGE Research Report, University of Lisbon.

Li, M., Ahmadiadli, Y., & Zhang, X.-P. (2025). A survey on speech deepfake detection. ACM Computing Surveys, 57(7), 1-38. https://doi.org/10.1145/3714458

Liu, X., Wang, X., Sahidullah, M., Patino, J., Delgado, H., Kinnunen, T., Todisco, M., Yamagishi, J., Evans, N., Nautsch, A., & Lee, K. A. (2023). ASVspoof 2021: Towards spoofed and deepfake speech detection in the wild. IEEE/ACM Transactions on Audio,

Speech, and Language Processing, 31, 2507-2522. https://doi.org/10.1109/TASLP.2023.3285283

Malinka, K., Firc, A., Kaška, P., Lapšanský, T., Šandor, O., & Homoliak, I. (2024). Resilience of voice assistants to synthetic speech. In J. Garcia-Alfaro, R. Kozik, M. Choraś, & S. Katsikas (Eds.), Computer Security - ESORICS 2024 (Lecture Notes in Computer Science, Vol. 14801, pp. 66-84). Springer. https://doi.org/10.1007/978-3-031-70879-4_4

Purdue University. (2025). PDID dataset benchmark: Evaluating commercial deepfake detection tools. Purdue University Research Report.

Reimao, R., & Tzerpos, V. (2019). FoR: A dataset for synthetic speech detection. 2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), 1-10. https://doi.org/10.1109/SPED.2019.8906599

Roy, D. (2025). Deepfake detection using convolutional neural networks: A literature survey. National Institute of Technology Karnataka Technical Report.

Sahidullah, M., Kinnunen, T., & Hanilçi, C. (2015). A comparison of features for synthetic speech detection. Interspeech 2015, 2087-2091. https://doi.org/10.21437/Interspeech.2015-472

Todisco, M., Wang, X., Vestman, V., Sahidullah, M., Delgado, H., Nautsch, A., Yamagishi, J., Evans, N., Kinnunen, T., & Lee, K. A. (2019). ASVspoof 2019: Future horizons in spoofed and fake audio detection. Interspeech 2019, 1008-1012. https://doi.org/10.21437/Interspeech.2019-2249

W3C. (2025). Web Content Accessibility Guidelines (WCAG) 3.0: Authentication and cognitive accessibility. W3C Working Draft.

Xie, Y., Zhang, Z., & Yang, Y. (2021). Siamese network with wav2vec feature for spoofing speech detection. Interspeech 2021, 4269-4273. https://doi.org/10.21437/Interspeech.2021-847

Xue, J., Fan, C., Lv, Z., Tao, J., Yi, J., Zheng, C., Wen, Z., Yuan, M., & Shao, S. (2022). Audio deepfake detection based on a combination of F0 information and real plus imaginary spectrogram features. Proceedings of the 30th ACM International Conference on Multimedia, 19-26. https://doi.org/10.1145/3552466.3556526

Yamazaki, Y., & Hasegawa, S. (2023). Navigation method enhancing music listening experience by stimulating both neck sides with modulated musical vibration. IEEE Transactions on Haptics, 16(2), 228-239. https://doi.org/10.1109/TOH.2023.3266194

Zhang, B., Cui, H., Nguyen, V., & Whitty, M. (2025). Audio deepfake detection: What has been achieved and what lies ahead. Sensors, 25(7), 1989. https://doi.org/10.3390/s25071989