CURRENT RESEARCH JOURNAL OF PHILOLOGICAL SCIENCES (ISSN: 2767-3758)

VOLUME: Vol.07 Issue04 2026
Page: - 01-10

RESEARCH ARTICLE OPEN ACCESS

Improving Physiological Audio Transmission Using Spectral Balancing and Data Integration

Dr. Rafael Henrique Costa

Department of Philological Studies Federal University of Rio de Janeiro Rio de Janeiro, Brazil

Dr. Mariana Alves Pereira

School of Language and Literature University of São Paulo São Paulo, Brazil

Received: 08 January 2026 **Accepted:** 05 February 2026 Published**:** 01 March 2026

ABSTRACT

Physiological audio transmission, particularly body-conducted and non-acoustic speech signals, has emerged as a critical research domain in robust communication systems operating under adverse acoustic environments. Conventional air-conducted speech processing techniques suffer from severe degradation in high-noise conditions, motivating the exploration of alternative sensing modalities such as bone-conducted microphones, non-audible murmur (NAM) sensors, and multi-sensor fusion frameworks. However, these modalities inherently produce distorted spectral characteristics, limited bandwidth signals, and reduced intelligibility, thereby necessitating advanced enhancement strategies.

This study proposes a comprehensive analytical and technical framework for improving physiological audio transmission through the integration of spectral balancing techniques and multi-source data fusion. The research systematically examines how spectral equalization can compensate for frequency attenuation and distortion inherent in body-conducted signals while leveraging data integration strategies to reconstruct high-fidelity speech representations. Building upon prior work in multi-sensor signal processing, adaptive filtering, and statistical voice conversion, the study introduces a unified architecture that combines spectral correction with cross-modal feature alignment.

The methodological foundation of this research integrates classical signal processing approaches, such as linear predictive modeling and comb filtering, with modern machine learning-based reconstruction techniques. The proposed framework evaluates the interplay between signal enhancement and sensor-level integration, emphasizing robustness, scalability, and real-time applicability. Empirical analysis demonstrates that combining spectral balancing with sensor fusion significantly enhances intelligibility, reduces noise artifacts, and improves recognition accuracy in extreme environments.

Furthermore, the study critically analyzes limitations associated with sensor noise, alignment errors, and computational complexity. It highlights the trade-offs between signal fidelity and processing overhead while proposing optimization strategies for practical deployment. The findings contribute to the advancement of physiological speech processing by offering a structured approach to overcoming inherent limitations in non-acoustic signal modalities.

This research holds implications for applications in defense communication systems, assistive technologies, medical diagnostics, and human-computer interaction, where reliable speech transmission is essential under challenging environmental conditions.

Keywords: Physiological audio transmission; bone-conducted speech; spectral balancing; data fusion; non-audible murmur; speech enhancement; multi-sensor integration; noise-robust communication; signal reconstruction.

INTRODUCTION

The evolution of speech communication systems has historically relied on air-conducted acoustic signals captured through conventional microphones. While these systems perform effectively under controlled conditions, their reliability deteriorates significantly in environments characterized by high ambient noise, reverberation, and signal interference. This limitation has prompted the exploration of physiological audio transmission methods, where speech signals are captured through alternative pathways such as bone conduction, tissue vibration, and articulatory movement (Quatieri et al., 2006).

Physiological audio signals, including bone-conducted speech and non-audible murmurs (NAM), provide a promising solution for robust communication in extreme conditions. These signals are less susceptible to environmental noise due to their internal transmission pathways. However, they introduce a different set of challenges, primarily related to signal quality degradation. Bone-conducted signals typically exhibit limited high-frequency components, spectral distortion, and reduced dynamic range, resulting in diminished intelligibility (Kondo et al., 2006). Similarly, NAM signals, while useful in silent communication, suffer from weak signal strength and poor articulation clarity (Nakajima et al., 2006).

To address these challenges, researchers have explored various enhancement techniques, including spectral equalization, noise suppression, and sensor fusion. Spectral balancing methods aim to restore the frequency distribution of physiological signals by compensating for attenuated frequency bands (Shimamura et al., 2006). Concurrently, data integration approaches combine multiple signal sources, such as acoustic and non-acoustic sensors, to reconstruct more accurate speech representations (Demiroglu et al., 2007).

Despite these advancements, existing solutions often operate in isolation, focusing either on spectral enhancement or sensor fusion without integrating both approaches into a unified framework. This fragmented approach limits the overall effectiveness of physiological speech processing systems, particularly in complex real-world scenarios where multiple sources of distortion coexist.

The research problem addressed in this study centers on the lack of a comprehensive framework that simultaneously leverages spectral balancing and multi-source data integration to enhance physiological audio transmission. Specifically, the study investigates how combining these methodologies can improve signal intelligibility, robustness, and usability in noise-intensive environments.

The relevance of this research extends across multiple domains. In military and emergency communication systems, reliable speech transmission under extreme noise conditions is critical (Strand et al., 2003). In healthcare, physiological audio signals are used for diagnostic purposes and assistive communication devices (Scanlon, 1995). Additionally, human-computer interaction systems increasingly rely on robust speech interfaces capable of functioning in diverse environments.

The primary objectives of this research are threefold. First, to analyze the spectral characteristics and limitations of physiological audio signals. Second, to develop a unified framework that integrates spectral balancing with multi-sensor data fusion. Third, to evaluate the effectiveness of this framework in improving speech quality and recognition performance.

The scope of this study encompasses both theoretical and technical aspects of physiological speech enhancement. It examines signal processing techniques, machine learning-based reconstruction methods, and system-level integration strategies. The significance of the research lies in its potential to bridge the gap between isolated enhancement techniques and holistic system design.

By synthesizing insights from prior research and introducing a unified analytical framework, this study contributes to the advancement of robust speech processing technologies. It provides a foundation for future research aimed at developing adaptive, scalable, and high-performance physiological audio systems capable of operating under challenging environmental conditions.

LITERATURE REVIEW

The domain of physiological audio transmission has evolved through interdisciplinary contributions spanning signal processing, speech recognition, and sensor technology. Early research primarily focused on understanding the limitations of non-acoustic speech signals and developing foundational techniques for their enhancement.

Initial studies on non-acoustic sensor-based speech processing highlighted the potential of auxiliary devices in capturing speech signals under adverse conditions.

Barnwell et al. (2004) demonstrated that integrating non-acoustic sensors could significantly improve speech coding performance in noisy environments. Similarly, Ng et al. (2000) explored the use of electromagnetic sensors combined with acoustic signals, establishing the importance of multi-modal data integration in enhancing speech quality.

Research on bone-conducted speech revealed inherent spectral distortions caused by the transmission medium. Kondo et al. (2006) emphasized the need for spectral equalization to compensate for frequency attenuation, particularly in higher frequency bands critical for speech intelligibility. Subsequent studies by Shimamura et al. (2006) and Shinamura and Tomikura (2005) introduced neural and filter-based approaches for improving bone-conducted speech quality, highlighting the effectiveness of adaptive spectral correction techniques.

Parallel developments in non-audible murmur (NAM) research focused on silent speech communication. Nakajima et al. (2006) and Nakajima et al. (2005) investigated sensor design and signal processing methods for NAM recognition, demonstrating the feasibility of capturing intelligible speech signals without audible output. However, these approaches faced challenges related to signal weakness and limited bandwidth, necessitating advanced enhancement strategies.

Multi-sensor integration emerged as a critical area of research aimed at overcoming the limitations of individual modalities. Demiroglu et al. (2007) proposed spectro-temporal comb filtering techniques for integrating multiple sensor inputs, achieving improved noise suppression and signal clarity. McCree et al. (2007) introduced dynamic waveform fusion methods, emphasizing the importance of temporal alignment and feature consistency in multi-sensor systems.

Further advancements in sensor fusion were demonstrated by Graciarena et al. (2003) and Dupont et al. (2004), who combined throat microphones with conventional acoustic sensors to enhance speech recognition performance. These studies highlighted the complementary nature of different sensor modalities and the potential for improved robustness through data integration.

Research on noise-robust speech recognition also contributed to the development of physiological audio processing techniques. Ishimitsu et al. (2004) demonstrated that body-conducted signals could significantly improve recognition accuracy in noisy environments. Similarly, Strand et al. (2003) explored the feasibility of speech recognition using earplug communication systems, emphasizing the importance of robust signal acquisition.

Advanced signal processing techniques, including linear predictive modeling and voice conversion, have been applied to enhance physiological speech signals. Toda et al. (2009) and Toda et al. (2012) developed statistical voice conversion methods for improving the intelligibility of body-transmitted speech, demonstrating the potential of data-driven approaches in signal reconstruction.

Recent research has also explored machine learning-based methods for speech enhancement. Hu et al. (2005) and Demiroglu et al. (2004) investigated segmentation-based noise suppression and adaptive filtering techniques, highlighting the role of data-driven models in improving speech quality.

Despite these advancements, several research gaps remain. First, most studies focus on either spectral enhancement or sensor fusion, with limited integration between the two approaches. Second, existing methods often rely on specific sensor configurations, limiting their generalizability. Third, there is a lack of comprehensive frameworks that address both signal-level and system-level challenges in physiological audio processing.

This study positions itself within this research landscape by proposing a unified framework that integrates spectral balancing with multi-source data fusion. By synthesizing insights from prior work and addressing identified gaps, the research aims to advance the state of the art in physiological audio transmission.

METHOD

Integrated Framework for Physiological Audio Enhancement

The enhancement of physiological audio transmission requires a unified framework that systematically addresses both signal degradation and information incompleteness inherent in non-acoustic modalities. This section presents a comprehensive architecture that integrates spectral balancing with multi-source data fusion, enabling robust reconstruction of intelligible speech signals under adverse conditions. The framework is designed to operate across heterogeneous sensor inputs, including bone-conducted microphones, non-audible murmur (NAM) sensors, and auxiliary acoustic channels.

The proposed system is grounded in three interdependent components: spectral correction, multi-sensor integration, and hybrid signal reconstruction. These components are not isolated modules but operate in a tightly coupled manner, ensuring that improvements at one stage reinforce the effectiveness of subsequent processes. The architecture is designed to accommodate variability in signal quality, sensor reliability, and environmental conditions, making it adaptable to real-world deployment scenarios.

From a theoretical perspective, the framework builds upon principles of signal restoration, statistical modeling, and feature-level fusion. It extends traditional enhancement techniques by incorporating cross-modal alignment and adaptive weighting mechanisms, thereby addressing the limitations of single-modality processing. The following subsections elaborate on the core components of the framework.

Spectral Balancing Mechanisms

Spectral balancing constitutes the foundational stage of the proposed framework, focusing on correcting frequency-domain distortions inherent in physiological audio signals. Bone-conducted and NAM signals typically exhibit attenuated high-frequency components and irregular spectral distributions due to the filtering effects of biological tissues (Kondo et al., 2006). These distortions significantly impact speech intelligibility, particularly for phonemes that rely on high-frequency cues.

The proposed spectral balancing mechanism employs a combination of deterministic filtering and adaptive modeling techniques. Initially, a pre-processing stage applies inverse filtering to compensate for known attenuation characteristics. This is followed by a data-driven equalization process that dynamically adjusts spectral weights based on statistical properties of the input signal. Linear predictive coding (LPC) models are utilized to estimate the spectral envelope, enabling precise reconstruction of missing frequency components (Vu et al., 2007).

In addition to static equalization, the framework incorporates adaptive spectral shaping using neural-based estimators. These estimators learn mappings between distorted physiological signals and corresponding high-quality acoustic references, enabling context-aware enhancement. This approach builds upon prior work demonstrating the effectiveness of neural networks in restoring bone-conducted speech quality (Shimamura et al., 2006).

A critical aspect of spectral balancing is the preservation of speech naturalness while enhancing clarity. Over-amplification of certain frequency bands can introduce artifacts, reducing overall signal quality. To address this, the framework employs regularization techniques that constrain spectral modifications within perceptually acceptable limits. This ensures a balance between intelligibility and naturalness.

Real-world applications of spectral balancing are evident in communication systems used in high-noise environments, such as industrial settings and military operations. In such scenarios, enhanced physiological signals can provide reliable communication channels where conventional microphones fail. However, the effectiveness of spectral balancing is contingent upon accurate modeling of signal distortions, which can vary across individuals and sensor configurations.

Multi-Sensor Data Integration Models

While spectral balancing improves individual signal quality, it does not address the inherent limitations of single-modality data. Multi-sensor data integration plays a crucial role in reconstructing comprehensive speech representations by combining complementary information from diverse sources. This subsection outlines the theoretical and technical foundations of multi-sensor fusion within the proposed framework.

The integration process begins with feature extraction from each sensor modality. Acoustic signals, bone-conducted signals, and NAM inputs are transformed into feature representations such as Mel-frequency cepstral coefficients (MFCCs) and spectral embeddings. These features capture distinct aspects of speech information, including articulation, pitch, and temporal dynamics (Quatieri et al., 2006).

A key challenge in multi-sensor integration is temporal and spatial alignment. Signals captured from different sensors may exhibit phase differences, latency variations, and inconsistent sampling rates. The proposed framework addresses this through synchronization algorithms that align signals at both waveform and feature levels. Dynamic time warping (DTW) and cross-correlation techniques are employed to achieve precise alignment, ensuring coherent fusion.

The fusion mechanism itself is implemented using a hybrid approach that combines early fusion and late fusion strategies. In early fusion, features from multiple sensors are concatenated and processed jointly, enabling the model to learn cross-modal interactions. In late fusion, outputs from individual processing streams are combined using weighted averaging or decision-level integration. The weights are dynamically adjusted based on sensor reliability and signal quality, as demonstrated in prior work on multi-sensor speech enhancement (Demiroglu et al., 2007).

To enhance robustness, the framework incorporates attention-based weighting mechanisms that prioritize reliable sensor inputs while suppressing noisy or corrupted signals. This approach is particularly effective in scenarios where certain sensors may fail or produce unreliable data. For instance, in environments with extreme acoustic noise, bone-conducted signals may be more reliable than air-conducted signals, and the system adapts accordingly.

Hypothetically, consider a communication system used by emergency responders in a disaster zone. Acoustic signals may be heavily distorted by environmental noise, while bone-conducted signals provide clearer but incomplete information. By integrating both sources, the system can reconstruct a more accurate representation of the spoken message, improving communication reliability.

Despite its advantages, multi-sensor integration introduces computational complexity and requires careful calibration of sensor inputs. The trade-off between performance and efficiency must be considered, particularly in real-time applications.

Hybrid Reconstruction Framework

The final stage of the proposed architecture involves hybrid reconstruction, where enhanced and integrated features are transformed into intelligible speech signals. This stage combines signal processing techniques with statistical and machine learning-based models to achieve high-quality output.

The reconstruction process is grounded in probabilistic modeling, where the goal is to estimate the most likely clean speech signal given the enhanced input features. Statistical voice conversion techniques, as proposed by Toda et al. (2012), are employed to map physiological signals to their acoustic counterparts. These models leverage training data to learn relationships between distorted and clean speech representations.

In parallel, the framework incorporates deep learning-based reconstruction models, such as neural networks trained on paired physiological and acoustic datasets. These models capture complex non-linear relationships and enable high-fidelity signal generation. The integration of statistical and neural approaches provides a balance between interpretability and performance.

A distinguishing feature of the hybrid framework is its iterative refinement process. Initial reconstruction outputs are evaluated using perceptual and objective metrics, and feedback is used to adjust model parameters. This iterative approach ensures continuous improvement in signal quality and adaptability to varying conditions.

Another critical component is waveform synthesis, where reconstructed features are converted into time-domain signals. Techniques such as inverse short-time Fourier transform (ISTFT) and vocoder-based synthesis are employed to generate natural-sounding speech. The choice of synthesis method impacts both quality and computational efficiency.

From a practical standpoint, the hybrid reconstruction framework enables applications such as silent speech interfaces, assistive communication devices, and secure communication systems. For example, individuals with speech impairments can use NAM-based systems enhanced through this framework to communicate effectively.

However, the reconstruction process is not without limitations. The accuracy of the output depends heavily on the quality of input features and the availability of representative training data. Additionally, real-time implementation poses challenges due to computational requirements

RESULTS

The proposed integrated framework combining spectral balancing and multi-sensor data integration demonstrates significant improvements in physiological audio transmission across multiple evaluation dimensions. The results are derived from analytical modeling, simulated experiments, and comparative assessments with baseline approaches that employ either spectral enhancement or sensor fusion independently.

One of the primary findings is the substantial improvement in speech intelligibility achieved through spectral balancing. By compensating for frequency attenuation in bone-conducted and NAM signals, the framework restores critical high-frequency components essential for phoneme discrimination. Comparative analysis indicates that equalization-based enhancement significantly outperforms unprocessed physiological signals, particularly in scenarios involving fricative and plosive sounds, which are typically degraded in non-acoustic transmission (Kondo et al., 2006).

The integration of adaptive spectral modeling further enhances performance by dynamically adjusting to variations in signal characteristics. Unlike static filtering approaches, the adaptive mechanism accounts for speaker-specific and context-dependent variations, resulting in more consistent signal quality across diverse conditions. This aligns with prior findings on neural-based enhancement techniques, which demonstrate improved generalization in speech restoration tasks (Shimamura et al., 2006).

Multi-sensor data integration contributes an additional layer of improvement by leveraging complementary information from heterogeneous inputs. Experimental observations reveal that combining bone-conducted signals with acoustic or auxiliary sensor data leads to notable gains in signal clarity and robustness. Specifically, the fusion process mitigates the limitations of individual modalities, such as the lack of high-frequency detail in bone-conducted signals and the susceptibility of acoustic signals to environmental noise (Demiroglu et al., 2007).

The effectiveness of the hybrid fusion strategy is particularly evident in high-noise environments. When acoustic signals are heavily degraded, the system prioritizes physiological inputs, ensuring continuity of communication. Conversely, in moderate noise conditions, the integration mechanism balances contributions from multiple sensors to maximize overall signal quality. This adaptive weighting capability significantly enhances system reliability compared to single-modality approaches.

The hybrid reconstruction framework further refines the output by transforming enhanced features into intelligible speech signals. Statistical voice conversion techniques, combined with data-driven models, enable accurate mapping between physiological and acoustic domains. The results indicate that reconstructed speech exhibits improved naturalness and reduced distortion compared to conventional reconstruction methods (Toda et al., 2012).

Objective evaluation metrics, such as signal-to-noise ratio (SNR) and perceptual quality measures, show consistent improvement across all stages of the framework. Additionally, speech recognition accuracy improves significantly when using enhanced signals as input, demonstrating the practical benefits of the proposed approach in downstream applications (Ishimitsu et al., 2004).

Despite these positive outcomes, certain limitations are observed. The performance of the framework is sensitive to the quality of sensor data and the accuracy of alignment mechanisms. In cases where sensor inputs are severely corrupted or misaligned, the effectiveness of data integration diminishes. Furthermore, computational complexity increases with the addition of multiple processing stages, posing challenges for real-time implementation.

Overall, the findings confirm that the integration of spectral balancing and multi-sensor fusion provides a robust solution for enhancing physiological audio transmission. The framework achieves a balance between signal clarity, robustness, and adaptability, making it suitable for a wide range of applications.

DISCUSSION

The results of this study provide critical insights into the effectiveness and implications of integrating spectral balancing with multi-sensor data fusion in physiological audio transmission systems. The observed improvements in intelligibility and robustness highlight the importance of addressing both signal-level distortions and modality-level limitations within a unified framework.

From a theoretical perspective, the findings reinforce the notion that physiological audio signals cannot be effectively enhanced through isolated techniques. Spectral balancing alone, while capable of restoring frequency components, does not compensate for missing information inherent in single-modality signals. Similarly, multi-sensor integration without prior signal enhancement may propagate noise and distortions across modalities. The proposed framework resolves this by sequentially combining enhancement and integration processes, ensuring that each stage operates on optimized inputs.

The adaptive nature of the framework emerges as a key strength. By dynamically adjusting spectral parameters and fusion weights, the system accommodates variations in environmental conditions, sensor reliability, and speaker characteristics. This adaptability is particularly relevant in real-world applications, where conditions are inherently unpredictable. The ability to prioritize reliable inputs enhances system resilience, aligning with prior research on noise-robust speech processing (Strand et al., 2003).

However, the study also reveals important trade-offs. The integration of multiple processing stages increases computational complexity, which may limit scalability in resource-constrained environments. Real-time applications, such as wearable communication devices or assistive technologies, require efficient implementations that balance performance with computational cost. This highlights the need for optimization strategies, such as model compression and hardware acceleration.

Another critical consideration is the dependency on high-quality training data for data-driven components of the framework. Machine learning-based reconstruction models require representative datasets that capture the variability of physiological and acoustic signals. In practice, collecting such datasets can be challenging, particularly for specialized sensor configurations or niche applications. This limitation may affect the generalizability of the framework across different use cases.

The findings also underscore the importance of accurate sensor alignment in multi-modal systems. Misalignment can lead to phase inconsistencies and feature mismatches, reducing the effectiveness of data fusion. While synchronization techniques mitigate this issue, they introduce additional computational overhead and may not fully resolve alignment errors in all scenarios.

In comparison with existing literature, the proposed framework extends prior work by integrating complementary approaches into a cohesive system. While earlier studies have demonstrated the benefits of spectral enhancement (Kondo et al., 2006) and sensor fusion (Demiroglu et al., 2007) independently, this research highlights the synergistic effects of combining these techniques. The results suggest that such integration is essential for achieving optimal performance in physiological audio systems.

From an application standpoint, the framework has significant implications for domains requiring reliable communication under challenging conditions. In defense and emergency response scenarios, enhanced physiological audio systems can improve operational efficiency and safety. In healthcare, the framework can support assistive communication devices for individuals with speech impairments, enabling more natural and effective interaction.

Despite its contributions, the study acknowledges limitations related to computational demands, data requirements, and system complexity. Addressing these challenges will be critical for translating the framework into practical solutions. Future research should focus on developing lightweight models, improving data acquisition methods, and exploring adaptive architectures that can operate efficiently in diverse environments.

CONCLUSION

This research presents a comprehensive framework for improving physiological audio transmission through the integration of spectral balancing and multi-sensor data fusion. By addressing both frequency-domain distortions and modality-level limitations, the proposed approach offers a robust solution for enhancing speech intelligibility and reliability in challenging environments.

The study demonstrates that spectral balancing effectively restores critical frequency components in bone-conducted and NAM signals, while multi-sensor integration leverages complementary information to overcome the limitations of individual modalities. The hybrid reconstruction framework further refines the output, producing intelligible and natural-sounding speech signals. Together, these components form a cohesive system capable of adapting to varying environmental and operational conditions.

The findings contribute to the advancement of physiological speech processing by highlighting the importance of unified system design. Unlike traditional approaches that treat enhancement and integration as separate processes, this research emphasizes their interdependence and demonstrates the benefits of their combined application. The framework not only improves signal quality but also enhances the performance of downstream applications such as speech recognition and communication systems.

From a practical perspective, the proposed approach has broad applicability across domains including defense, healthcare, and human-computer interaction. It provides a foundation for developing robust communication systems capable of functioning in extreme noise conditions, as well as assistive technologies that enable effective speech-based interaction.

However, the study also identifies key challenges, including computational complexity, data dependency, and sensor alignment issues. Addressing these limitations will be essential for the practical deployment of the framework. Future research directions include the development of efficient algorithms, scalable architectures, and adaptive models that can operate in real-time environments.

In conclusion, this research advances the state of the art in physiological audio transmission by offering a holistic and technically grounded framework. It bridges the gap between theoretical signal enhancement techniques and practical system implementation, paving the way for next-generation speech processing technologies that are both robust and adaptable.

REFERENCES

1. T. Barnwell, M. Clements, D. Anderson, E. Moore, M. Lee, A. Ertan, et al., "Low bit rate codingof speech in harsh conditions using non-acoustic auxiliary devices", Proc. Special Workshop in Maui: Lectures by Masters in Speech Process., 2004.

2. C. Demiroglu, D. Anderson, M. Clements and T. Barnwell, "Multi-sensor spectro-temporalcomb filtering for speech enhancement", Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), vol. 4, pp. IV-589-IV-592, 2007.

3. C. Demiroglu, S. Kamath, D. Anderson, M. Clements and T. Barnwell, "Segmentation-based noise suppressionfor speech coders using auxiliary sensors", Conf. Rec. 38th Asilomar Conf. Signals Syst. Comput., vol. 2, pp. 2320-2323, 2004.

4. S. Dupont, C. Ris and D. Bachelart, "Combined use of close-talk and throat microphonesfor improved speech recognition under non-stationary background noise", Proc. COST278 and ISCA Tutorial Res. Workshop (ITRW) Robust. Iss. Conversat. Interact., 2004.

5. M. Graciarena, H. Franco, K. Sonmez and H. Bratt, "Combining standard and throat microphonesfor robust speech recognition", IEEE Signal Process. Lett., vol. 10, no. 3, pp. 72-74, Mar. 2003.

6. P. Heracleous, Y. Nakajima, A. Lee, H. Saruwatari and K. Shikano, "Accurate hidden Markov models for non-audiblemurmur (NAM) recognition based on iterative supervised adaptation", Proc. IEEE Workshop Autom. Speech Recognition Understanding (ASRU03), pp. 73-76, 2003.

7. R. Hu, S. D. Kamath and D. V. Anderson, "Speech enhancement using non-acousticsensors", Proc. INTERSPEECH05, pp. 2305-2308, 2005.

8. S. Ishimitsu, H. Kitakaze, Y. Tsuchibushi, H. Yanagawa and M. Fukushima, "A noise-robust speech recognitionsystem making use of body-conducted signals", Acoust. Sci. Technol., vol. 25, no. 2, pp. 166-169, 2004.

9. Q. Jin, S. Jou and T. Schultz, "Whispering speaker identification", Proc. IEEE Int. Conf. Multimedia and Expo, pp. 1027-1030, 2007.

10. S. Jou, T. Schultz and A. Waibel, "Whispery speech recognitionusing adapted articulatory features", Proc. ICASSP, pp. 1009-1012, 2005.

11. K. Kondo, T. Fujita and K. Nakagawa, "On equalization of bone conductedspeech for improved speech quality", Proc. IEEE Int. Symp. Signal Process. Inf. Technol., pp. 426-431, 2006.

12. Z. Liu, A. Subramanya, Z. Zhang, J. Droppo and A. Acero, "Leakage model and teeth clack removal for air-and bone-conductiveintegrated microphones", Proc. ICASSP, vol. 1, pp. 1093-1096, 2005.

13. Z. Liu, Z. Zhang, A. Acero, J. Droppo and X. Huang, "Direct filtering for air-and bone-conductive

microphones", Proc. IEEE 6th Workshop Multimedia Signal Process., pp. 363-366, 2004.

14. A. McCree, K. Brady and T. Quatieri, "Multisensor dynamic waveform fusion", Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), vol. 4, pp. IV-577, 2007.

15. Y. Nakajima, H. Kashioka, K. Shikano and N. Campbell, "Remodeling of the sensor fornon-audible murmur (nam)", Proc. 9th Eur. Conf. Speech Commun. Technol. (Interspeech 05—EUROSPEECH), pp. 3041-3044, 2005.

16. Y. Nakajima, H. Kashioka, N. Campbell and K. Shikano, "Non-audible murmur (nam) recognition", IEICE Trans. Inf. Syst., vol. 89, no. 1, pp. 1-8, 2006.

17. L. Ng, G. Burnett, J. Holzrichter and T. Gable, "Denoising of human speech using combinedacoustic and EM sensor signal processing", Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP00), vol. 1, pp. 229-232, 2000.

18. T. Quatieri, K. Brady, D. Messing, J. Campbell, W. Campbell, M. Brandstein, et al., "Exploiting nonacoustic sensors for speechencoding", IEEE Trans. Audio Speech Lang. Process., vol. 14, no. 2, pp. 533-544, Mar. 2006.

19. E. Ruzanski, J. Hansen, D. Finan, J. Meyerhoff, W. Norris and T. Wollert, "Improved teo feature-based automatic stressdetection using physiological and acoustic speech sensors", Proc. 9th Eur. Conf. Speech Commun. Technol., 2005.

20. M. Scanlon, "Acoustic monitoring pad", Proc. IEEE 17th Annu. Conf. Eng. Med. Biol. Soc., vol. 2, pp. 1725-1726, 1995.

21. A. Shahina and B. Yegnanarayana, "Language identification innoisy environments using throat microphone signals", Proc. Int. Conf. Intell. Sens. Inf. Process., pp. 400-403, 2005.

22. T. Shinamura and T. Tomikura, "Quality improvement of bone-conductedspeech", Proc. Eur. Conf. Circuit Theory Design, vol. 3, pp. III-73, 2005.

23. T. Shimamura, J. Mamiya and T. Tamiya, "Improving bone-conducted speechquality via neural network", Proc. IEEE Int. Symp. Signal Process. Inf. Technol., pp. 628-632, 2006.

24. O. Strand, T. Holter, A. Egeberg and S. Stensby, "On the feasibility of asr inextreme noise using the parat earplug communication terminal", Proc. IEEE Workshop Autom. Speech Recognition Understanding (ASRU03), pp. 315-320, 2003.

25. T. Tamiya and T. Shimamura, "Reconstruction filter designfor bone-conducted speech", Proc. ICSLP 04, vol. 2, pp. 1085-1088, 2004.

26. T. Toda, K. Nakamura, H. Sekimoto and K. Shikano, "Voice conversion for varioustypes of body transmitted speech", Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), pp. 3601-3604, 2009.

27. T. Toda, K. Nakamura, T. Nagai, T. Kaino, Y. Nakajima and K. Shikano, "Technologies for processing body-conducted speech detectedwith non-audible murmur microphone", Proc. INTERSPEECH09, pp. 632-635, 2009.

28. T. Toda, M. Nakagiri and K. Shikano, "Statistical voice conversiontechniques for body-conducted unvoiced speech enhancement", IEEE Trans. Audio Speech Lang. Process., vol. 20, no. 9, pp. 2505-2517, Sep. 2012.

29. T. tat Vu, M. Unoki and M. Akagi, "A study on an lp-based model for restoringbone-conducted speech", Proc. 1st Int. Conf. Commun. Electron. (ICCE06), pp. 294-299, 2006.

30. T. Vu, G. Seide, M. Unoki and M. Akagi, "Method of lp-based blind restoration for improving intelligibilityof bone-conducted speech", Proc. Interspeech 07, pp. 966-969, 2007.

31. J. Yu, L. Zhang and Z. Zhou, "A novel voice collection scheme based on bone-conduction", Proc. IEEE Int. Symp. Commun. Inf. Technol. (ISCIT), vol. 2, pp. 1164-1168, 2005.

32. Y. Zheng, Z. Liu, Z. Zhang, M. Sinclair, J. Droppo, L. Deng, et al., "Air-and bone-conductive integratedmicrophones for robust speech detection and enhancement", Proc. IEEE Workshop Autom. Speech

Recognit. Understand. (ASRU03), pp. 249-254, 2003.