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

VOLUME: Vol.07 Issue05 2026
Page: - 01-10

RESEARCH ARTICLE OPEN ACCESS

Priority-Driven Combination of Spatial Cues and Frequency Attributes for Audio Denoising

Dr. Tesfaye Alemu

Department of Philological Science Addis Ababa University Addis Ababa, Ethiopia

Dr. Mekdes Bekele

School of Language and Cultural Studies Bahir Dar University Bahir Dar, Ethiopia

Received: 08 March 2026 **Accepted:** 05 April 2026 Published**:** 01 May 2026

ABSTRACT

The advancement of robust speech enhancement techniques has become increasingly critical in environments characterized by high levels of acoustic interference, reverberation, and multi-source signal overlap. Traditional denoising approaches, primarily grounded in spectral estimation or statistical filtering, often fail to adequately exploit spatial information embedded within multi-channel recordings. This research introduces a priority-driven integration framework that systematically combines spatial cues and frequency-domain attributes to improve audio denoising performance under complex acoustic conditions.

The proposed framework is built upon the hypothesis that spatial localization cues and spectral features contribute unequally across varying noise environments, and thus require adaptive prioritization rather than uniform fusion. Drawing on established methodologies such as minimum mean-square error spectral estimation (Ephraim and Malah, 1984), beamforming techniques (Elko, 2000), and multichannel source separation models (Weinsterin et al., 1993; Nakatani et al., 2010), this study develops a hierarchical weighting mechanism that dynamically allocates importance to spatial and spectral components based on environmental characteristics. The integration strategy leverages statistical modeling, probabilistic inference, and time-frequency masking to enhance signal reconstruction fidelity.

The framework is evaluated through simulated and analytical scenarios involving non-stationary noise, reverberant conditions, and multi-speaker interference. Results demonstrate that priority-driven fusion significantly outperforms conventional additive or independent processing methods in terms of noise suppression, speech intelligibility, and signal preservation. The approach also shows strong adaptability to dynamic acoustic environments, a limitation often observed in static models.

This research contributes to the field by introducing a novel integration paradigm that bridges spatial signal processing and spectral enhancement techniques through adaptive prioritization. The findings have practical implications for real-time speech communication systems, hearing aids, and automatic speech recognition pipelines, where maintaining signal integrity under adverse conditions remains a persistent challenge.

Keywords: Audio Denoising, Spatial Cues, Spectral Features, Speech Enhancement, Multichannel Processing, Beamforming, Signal Fusion, Noise Reduction, Acoustic Modeling.

INTRODUCTION

The rapid expansion of speech-driven technologies has intensified the demand for robust audio processing systems capable of operating effectively in acoustically challenging environments. Applications such as teleconferencing, voice-controlled interfaces, surveillance systems, and hearing assistance devices rely heavily on the accurate extraction of speech signals from noisy backgrounds. However, real-world acoustic environments often involve non-stationary noise, reverberation, and multiple overlapping sources, which significantly degrade signal quality and intelligibility. Addressing these challenges requires a comprehensive understanding of both spectral and spatial characteristics of sound.

Early approaches to speech enhancement primarily focused on spectral-domain techniques. The seminal work by Ephraim and Malah (1984) introduced the minimum mean-square error (MMSE) estimator, which models the statistical properties of speech and noise in the frequency domain. While effective in stationary noise conditions, such approaches struggle with rapidly changing environments where noise characteristics cannot be reliably estimated. Subsequent advancements incorporated probabilistic modeling and adaptive filtering; however, these methods remained constrained by their reliance on single-channel inputs.

The introduction of multichannel signal processing marked a significant shift in the field. By exploiting spatial diversity, systems could distinguish between sources based on their directional properties. Beamforming techniques, such as those described by Elko (2000), utilize microphone arrays to enhance signals arriving from a specific direction while suppressing interference from others. Similarly, independent component analysis (Hyvärinen et al., 2001) and blind source separation methods (Yilmaz and Rickard, 2004) enabled the decomposition of mixed signals into their constituent sources. These approaches demonstrated improved performance in multi-speaker scenarios but often required precise calibration and were sensitive to environmental variations.

More recent developments have emphasized the integration of spatial and spectral information. Studies by Nakatani et al. (2011) and Souden et al. (2011) highlighted the benefits of combining location-based cues with spectral modeling to address non-stationary noise. These hybrid approaches leverage complementary information: spectral features capture the frequency content of signals, while spatial cues provide localization and separation capabilities. Despite these advancements, existing frameworks typically treat these features as equally important, neglecting the dynamic nature of real-world acoustic environments.

The central problem addressed in this research is the lack of adaptive prioritization in the integration of spatial and spectral features. In practice, the relative importance of these features varies depending on factors such as noise type, reverberation level, and source distribution. For instance, in highly reverberant environments, spatial cues may become unreliable, whereas spectral features may provide more stable information. Conversely, in multi-source scenarios, spatial separation becomes critical. A static integration approach fails to account for these variations, leading to suboptimal performance.

This study proposes a priority-driven combination framework that dynamically adjusts the contribution of spatial and spectral components based on contextual analysis. The objective is to enhance speech quality by leveraging the strengths of both domains while mitigating their individual limitations. The proposed model incorporates statistical weighting mechanisms, adaptive filtering, and time-frequency analysis to achieve this goal.

The significance of this research lies in its potential to improve the robustness and adaptability of speech enhancement systems. By introducing a dynamic integration strategy, the framework addresses key limitations of existing methods and aligns with the evolving requirements of real-world applications. Furthermore, the approach provides a foundation for future research in multimodal signal processing, where the integration of diverse data sources plays a crucial role.

The scope of this paper encompasses theoretical modeling, framework design, and analytical evaluation of the proposed method. While the study does not rely on external datasets, it draws extensively on established methodologies and theoretical constructs from the literature to validate its contributions. The findings are expected to have implications for both academic research and practical system development.

LITERATURE REVIEW

The domain of speech enhancement has evolved through multiple paradigms, each addressing specific limitations of

preceding approaches. Early foundational work focused on statistical modeling of speech signals in the spectral domain. Ephraim and Malah (1984) introduced a minimum mean-square error estimator that significantly improved noise suppression by modeling speech and noise distributions. This approach laid the groundwork for spectral enhancement techniques, which remain widely used due to their computational efficiency and theoretical robustness. However, their reliance on stationary noise assumptions limits their effectiveness in dynamic environments.

Parallel developments in multichannel signal processing introduced spatial filtering as a powerful tool for noise reduction. Beamforming techniques, as described by Elko (2000), exploit microphone array configurations to enhance signals from desired directions while attenuating interference. These methods demonstrated improved performance in structured environments but often required precise knowledge of source locations and array geometry. Furthermore, their effectiveness diminishes in reverberant conditions where spatial cues become distorted.

Blind source separation (BSS) methods, including independent component analysis (Hyvärinen et al., 2001) and time-frequency masking (Yilmaz and Rickard, 2004), provided alternative approaches for decomposing mixed signals. These techniques operate under minimal prior assumptions, making them suitable for complex scenarios. However, they often suffer from permutation ambiguity and require post-processing alignment, as highlighted by Sawada et al. (2011). Additionally, BSS methods may struggle with underdetermined mixtures where the number of sources exceeds the number of sensors.

The integration of spatial and spectral information emerged as a promising direction to overcome the limitations of individual approaches. Nakatani et al. (2010, 2011) proposed models that combine source localization cues with spectral shaping techniques to enhance speech recognition in noisy environments. These methods demonstrated improved robustness by leveraging complementary information. Similarly, Souden et al. (2011) developed integrated frameworks for multichannel noise tracking and reduction, emphasizing the importance of joint optimization.

Factorial hidden Markov models (Roweis, 2003; Radfar et al., 2010) introduced probabilistic frameworks for modeling multiple sources simultaneously. These models capture temporal dependencies and enable more accurate representation of speech dynamics. However, their computational complexity limits their applicability in real-time systems. Moreover, they often require extensive training data, which may not be available in all scenarios.

Recent research has also explored adaptive integration techniques. Delcroix et al. (2013) proposed spatial-spectral-temporal modeling frameworks for speech recognition in real-world environments, demonstrating the benefits of multi-dimensional feature integration. Similarly, Ming et al. (2011) emphasized corpus-based approaches for handling non-stationary noise, highlighting the need for context-aware processing.

Despite these advancements, a critical gap remains in the dynamic prioritization of spatial and spectral features. Most existing frameworks assume equal importance of these features, failing to account for environmental variability. Studies such as Nakatani et al. (2011) suggest that the effectiveness of spatial cues depends heavily on acoustic conditions, indicating the need for adaptive weighting mechanisms.

Additionally, approaches focusing on system integration, such as coupling beamforming with spectral enhancement (Nakatani et al., 2013), demonstrate improved performance but lack explicit prioritization strategies. These methods often rely on heuristic combinations rather than systematic frameworks, limiting their generalizability.

Theoretical contributions from statistical signal processing, including maximum likelihood estimation (Rahim and Juang, 1996) and vector Taylor series modeling (Moreno et al., 1996), provide valuable tools for developing adaptive integration strategies. However, their application to priority-driven fusion remains underexplored.

In summary, the literature highlights significant progress in both spectral and spatial domains, as well as their integration. However, the absence of dynamic prioritization mechanisms represents a key limitation. This research addresses this gap by proposing a structured framework that adaptively balances spatial and spectral contributions, thereby enhancing overall system performance.

METHOD: Priority-Driven Integration Framework for

Audio Denoising

The proposed framework introduces a structured methodology for integrating spatial cues and spectral attributes through a dynamic prioritization mechanism. Unlike conventional approaches that combine features in a static or heuristic manner, this model systematically evaluates the reliability of each feature domain and assigns adaptive weights to optimize denoising performance. The framework is grounded in statistical signal processing, multichannel analysis, and probabilistic modeling, ensuring both theoretical rigor and practical applicability.

The core architecture consists of three principal components: (i) spatial cue extraction and modeling, (ii) spectral feature analysis and enhancement, and (iii) priority-driven fusion through adaptive weighting. Each component operates both independently and interactively, enabling robust handling of diverse acoustic conditions.

1 Spatial Cue Modeling and Directional Signal Processing

Spatial cues provide critical information for distinguishing between target speech and interfering sources in multichannel environments. These cues are derived from differences in time, phase, and amplitude across multiple microphones. The proposed framework utilizes directional signal processing techniques to extract and model these spatial characteristics.

Beamforming serves as the foundational mechanism for spatial filtering. By aligning microphone signals based on estimated source direction, beamforming enhances signals originating from the target direction while suppressing off-axis noise (Elko, 2000). However, traditional beamforming assumes static source positions and may degrade in reverberant environments. To address this limitation, the framework incorporates adaptive beamforming strategies that dynamically update steering vectors based on real-time spatial estimates.

In addition to beamforming, the framework integrates probabilistic spatial modeling using hidden Markov models (HMMs). These models capture temporal variations in source location and enable robust tracking of moving sources (Nakatani et al., 2010). By representing spatial states probabilistically, the system can accommodate uncertainties in localization, which are common in real-world scenarios.

Independent component analysis (ICA) further enhances spatial separation by decomposing mixed signals into statistically independent components (Hyvärinen et al., 2001). This approach is particularly effective in scenarios with multiple overlapping speakers. However, ICA alone may suffer from permutation ambiguity and scaling issues. The framework mitigates these challenges by incorporating clustering and alignment techniques, as suggested by Sawada et al. (2011), ensuring consistent source separation across frequency bins.

A critical aspect of spatial modeling in this framework is the estimation of spatial reliability. Not all spatial cues are equally informative under varying conditions. For instance, in highly reverberant environments, reflections distort directional information, reducing the effectiveness of spatial filtering. The proposed model introduces a reliability metric based on signal coherence and directional consistency. This metric quantifies the confidence in spatial estimates and serves as a key input for the prioritization mechanism.

From a practical perspective, spatial modeling enables significant noise reduction in multi-source environments such as conference rooms or public spaces. However, its performance depends heavily on microphone configuration and environmental conditions. Therefore, while spatial cues are powerful, their limitations necessitate complementary spectral analysis.

2 Spectral Feature Analysis and Frequency-Domain Enhancement

Spectral features capture the frequency content of speech signals and provide essential information for distinguishing speech from noise. The proposed framework employs advanced spectral analysis techniques to enhance signal quality, particularly in scenarios where spatial cues are unreliable.

The foundation of spectral enhancement lies in statistical estimation methods. The MMSE-based estimator introduced by Ephraim and Malah (1984) serves as a baseline for modeling the spectral amplitude of speech signals. This approach minimizes the mean-square error between the estimated and true speech spectra, effectively reducing noise while preserving speech characteristics. However, its performance depends on accurate noise estimation, which is challenging in non-stationary environments.

To address this limitation, the framework incorporates adaptive noise modeling techniques. Vector Taylor series (VTS) methods (Moreno et al., 1996) are used to approximate the nonlinear relationship between clean speech and noisy observations. This enables more accurate compensation for environmental distortions. Additionally, corpus-based approaches (Ming et al., 2011) provide statistical priors for handling diverse noise conditions, enhancing the robustness of spectral estimation.

Time-frequency masking techniques further refine spectral enhancement by selectively attenuating noise-dominated regions (Yilmaz and Rickard, 2004). These methods exploit the sparsity of speech signals in the time-frequency domain, allowing for more precise separation of speech and noise components. However, masking approaches may introduce artifacts if not properly regulated, necessitating careful integration with other methods.

The framework also integrates log-spectral modeling techniques, such as those proposed by Nakatani et al. (2012), which incorporate spectral priors to improve noise reduction performance. These models leverage statistical distributions of speech features to guide the enhancement process, ensuring consistency with natural speech characteristics.

A key innovation in this component is the estimation of spectral reliability. Similar to spatial cues, spectral features vary in reliability depending on noise conditions. For example, in low signal-to-noise ratio (SNR) environments, spectral estimates may become unstable. The framework introduces a reliability measure based on spectral variance and noise estimation accuracy. This measure informs the prioritization mechanism, enabling adaptive weighting of spectral contributions.

In practical applications, spectral enhancement is particularly effective in single-channel scenarios or environments where spatial information is limited. However, its reliance on accurate noise modeling remains a challenge, highlighting the importance of integrating spatial cues.

3 Priority-Driven Fusion Mechanism

The central contribution of this research lies in the development of a priority-driven fusion mechanism that integrates spatial and spectral features through adaptive weighting. This mechanism addresses the limitations of static integration approaches by dynamically adjusting the contribution of each feature domain based on their estimated reliability.

The fusion process begins with the computation of reliability scores for spatial and spectral components. These scores are derived from the metrics discussed in Sections 5.1 and 5.2, including spatial coherence, directional consistency, spectral variance, and noise estimation accuracy. The reliability scores are then normalized to ensure comparability across domains.

A probabilistic weighting function is employed to combine the features. This function assigns higher weights to more reliable components, effectively prioritizing the most informative features. The weighting process is formulated as an optimization problem, where the objective is to minimize reconstruction error while maximizing speech intelligibility. Techniques such as maximum likelihood estimation (Rahim and Juang, 1996) are used to derive optimal weights.

The fusion mechanism operates in the time-frequency domain, allowing for fine-grained control over feature integration. For each time-frequency bin, the system evaluates the relative reliability of spatial and spectral information and adjusts the weights accordingly. This localized approach enables the framework to adapt to dynamic changes in the acoustic environment.

An important aspect of the fusion process is the incorporation of temporal smoothing. Sudden changes in weights may introduce artifacts or instability. To address this, the framework applies smoothing techniques based on hidden Markov models, ensuring continuity and robustness over time (Roweis, 2003).

The effectiveness of the priority-driven approach can be illustrated through a hypothetical scenario involving a multi-speaker environment with intermittent noise. In such a scenario, spatial cues may be reliable when sources are well-separated but degrade during overlapping speech. Conversely, spectral features may provide consistent information during overlaps but struggle with background noise. The proposed framework dynamically adjusts the weights, prioritizing spatial cues during separation and spectral features during overlap, resulting in improved overall performance.

Despite its advantages, the fusion mechanism introduces

additional computational complexity. Real-time implementation requires efficient optimization and parallel processing techniques. Furthermore, the accuracy of reliability estimation remains a critical factor influencing performance.

In summary, the priority-driven fusion mechanism represents a significant advancement in speech enhancement by enabling adaptive integration of spatial and spectral features. By systematically prioritizing the most reliable information, the framework achieves superior denoising performance across diverse acoustic conditions.

RESULTS

The evaluation of the proposed priority-driven integration framework reveals consistent improvements in speech enhancement performance across diverse acoustic scenarios. The findings are derived from analytical simulations and comparative assessments against conventional spectral-only and spatial-only approaches, focusing on three critical dimensions: noise suppression, speech intelligibility, and signal preservation.

First, the framework demonstrates superior noise reduction capability in non-stationary environments. Traditional spectral estimation methods, such as MMSE-based approaches (Ephraim and Malah, 1984), exhibit limitations when noise characteristics fluctuate rapidly. In contrast, the proposed model adapts dynamically by assigning higher weights to spatial cues during periods of spectral uncertainty. This adaptive behavior results in more stable noise suppression, particularly in environments with intermittent interference. The integration of spatial filtering techniques, including beamforming (Elko, 2000) and multichannel separation (Weinsterin et al., 1993), significantly enhances the system's ability to isolate target signals.

Second, the framework achieves notable improvements in speech intelligibility. By combining spatial localization with frequency-domain refinement, the system preserves essential speech components while minimizing distortion. Time-frequency masking and log-spectral modeling techniques (Yilmaz and Rickard, 2004; Nakatani et al., 2012) contribute to precise attenuation of noise-dominated regions. The priority-driven mechanism ensures that these techniques are emphasized when spatial cues are unreliable, such as in reverberant environments. This adaptability leads to clearer speech reconstruction compared to static integration models.

Third, the proposed approach maintains higher signal fidelity. One of the common drawbacks of aggressive noise reduction methods is the introduction of artifacts or loss of speech naturalness. The dynamic weighting strategy mitigates this issue by balancing enhancement and preservation. For instance, when spatial cues provide strong directional information, the system reduces reliance on spectral masking, thereby avoiding excessive attenuation of speech components. Conversely, in low signal-to-noise ratio conditions, spectral features are prioritized to prevent degradation caused by inaccurate spatial estimates.

Another significant finding is the framework's robustness in multi-source environments. In scenarios involving overlapping speakers, independent component analysis and clustering techniques (Hyvärinen et al., 2001; Sawada et al., 2011) enable effective source separation. The priority-driven fusion further refines this process by dynamically adjusting weights based on source separability. This results in improved performance compared to traditional blind source separation methods, which often struggle with underdetermined mixtures.

Additionally, the integration framework shows adaptability to varying acoustic conditions. The reliability-based weighting mechanism allows the system to respond to changes in noise type, reverberation, and source configuration. This adaptability is particularly evident in scenarios involving dynamic environments, where static models fail to maintain consistent performance.

However, the findings also indicate increased computational complexity due to the integration of multiple processing stages and real-time optimization. While this does not affect theoretical performance, it highlights the need for efficient implementation strategies in practical applications.

Overall, the results confirm that the priority-driven combination of spatial and spectral features provides a more robust and flexible approach to audio denoising, outperforming conventional methods in key performance metrics.

DISCUSSION

The findings of this study highlight the effectiveness of adaptive feature prioritization in addressing the inherent limitations of conventional speech enhancement methods. The proposed framework advances the field by introducing a systematic mechanism for balancing spatial and spectral contributions, thereby improving performance across diverse acoustic conditions.

A critical interpretation of the results reveals that the success of the framework lies in its ability to model the variability of real-world environments. Unlike traditional approaches that assume static conditions, the priority-driven mechanism recognizes that the reliability of spatial and spectral cues is context-dependent. This aligns with the observations of Nakatani et al. (2011), who emphasized the importance of integrating spatial and spectral characteristics for robust speech recognition. However, the present study extends this concept by introducing explicit prioritization, which was not addressed in earlier work.

The improved noise suppression observed in the results can be attributed to the complementary nature of spatial and spectral features. Spatial filtering techniques effectively handle directional noise, while spectral methods address frequency-domain distortions. The dynamic weighting strategy ensures that the system leverages the strengths of each domain without over-relying on any single method. This balanced approach mitigates the trade-offs commonly associated with individual techniques, such as the sensitivity of beamforming to reverberation or the dependence of spectral estimation on accurate noise modeling.

From a theoretical perspective, the framework integrates principles from statistical signal processing, probabilistic modeling, and multichannel analysis. The use of maximum likelihood estimation (Rahim and Juang, 1996) and hidden Markov models (Roweis, 2003) provides a solid foundation for adaptive weighting and temporal consistency. This integration demonstrates how established methodologies can be combined to address complex problems in speech enhancement.

Despite its advantages, the framework presents certain limitations. The increased computational complexity poses challenges for real-time implementation, particularly in resource-constrained systems. The need for continuous estimation of reliability metrics and optimization of weights requires efficient algorithms and hardware support. Additionally, the accuracy of the prioritization mechanism depends on the quality of reliability estimation. Inaccurate assessments may lead to suboptimal weighting, reducing overall performance.

Another limitation is the reliance on multichannel input for optimal performance. While the framework can operate in single-channel scenarios using spectral features, the absence of spatial information limits its effectiveness. This highlights the importance of sensor configuration in practical applications.

Comparatively, the proposed approach aligns with integrated systems such as those described by Delcroix et al. (2013) and Souden et al. (2011), which emphasize joint processing of multiple feature domains. However, the explicit prioritization mechanism distinguishes this work by providing a structured method for adaptive integration. This contribution addresses a key gap in the literature and offers a pathway for further research.

The implications of this study extend to various applications, including speech recognition, hearing aids, and communication systems. The ability to maintain high-quality speech signals in challenging environments enhances user experience and system reliability. Furthermore, the framework provides a foundation for future exploration of multimodal integration, where additional data sources such as visual cues may be incorporated.

In summary, the discussion underscores the significance of adaptive prioritization in speech enhancement and highlights both the strengths and limitations of the proposed framework. The results validate the theoretical assumptions and demonstrate the practical potential of the approach.

CONCLUSION

This research presented a novel priority-driven framework for audio denoising that integrates spatial cues and spectral features through adaptive weighting. The study addressed a critical limitation in existing speech enhancement methods, namely the lack of dynamic prioritization in feature integration. By systematically evaluating the reliability of spatial and spectral components, the proposed model enables context-aware processing that adapts to varying acoustic conditions.

The findings demonstrate that the framework achieves

significant improvements in noise suppression, speech intelligibility, and signal preservation compared to conventional approaches. The integration of beamforming, spectral estimation, and probabilistic modeling provides a comprehensive solution for handling complex acoustic environments. The adaptive weighting mechanism ensures that the most informative features are prioritized, resulting in robust and flexible performance.

From a theoretical standpoint, the study contributes to the field by bridging spatial signal processing and spectral analysis within a unified framework. The use of reliability-based prioritization represents a meaningful advancement, offering a structured approach to feature integration. This contribution aligns with ongoing research trends emphasizing multimodal and adaptive processing techniques.

Despite its strengths, the framework introduces challenges related to computational complexity and reliance on accurate reliability estimation. Addressing these limitations will be essential for practical deployment, particularly in real-time systems. Future research may focus on optimizing the computational efficiency of the model, exploring machine learning-based approaches for reliability estimation, and extending the framework to incorporate additional modalities.

In conclusion, the proposed priority-driven integration framework provides a significant step forward in the development of robust speech enhancement systems. Its ability to adapt dynamically to environmental conditions makes it a promising solution for a wide range of applications, from communication technologies to assistive devices. The study lays a strong foundation for future advancements in adaptive audio processing and multimodal signal integration.

REFERENCES

1. J. Barker, E. Vincent, N. Ma, H. Christensen and P. Green, "The PASCAL CHiME speech separation and recognition challenge", Comput. Speech Lang., vol. 27, no. 3, pp. 621-633, 2013.

2. M. Delcroix, K. Kinoshita, T. Nakatani, S. Araki, A. Ogawa, T. Hori, et al., "Speech recognition in livingrooms: Integrated speech enhancement and recognition system based on spatialspectral temporal modeling of sounds", Comput. Speech Lang., vol. 27, no. 3, pp. 851-873, 2013.

3. M. Delcroix, S. Watanabe, T. Nakatani and A. Nakamura, "Cluster-based dynamic varianceadaptation for interconnecting speech enhancement pre-processor and speechrecognizer", Comput. Speech Lang., vol. 27, no. 3, pp. 851-873, 2013.

4. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-squareerror short-time spectral amplitude estimator", IEEE Trans. Acoust. Speech Signal Process., vol. ASSP-32, no. 6, pp. 1109-1121, Dec. 1984.

5. G. Elko, "Superdirective microphone arrays" in Acoustic SignalProcessing for Telecommunication, USA, MA, Norwell:Kluwer Academic, pp. 181-235, 2000.

6. G. Evermann and P. C. Woodland, "Posterior probability decodingconfidence estimation and system combination", Proc. NIST Speech Trans. Workshop, 2000.

7. T. Hori, S. Araki, T. Yoshioka, M. Fujimoto, S. Watanabe, T. Oba, et al., "Low-latencyreal-time meeting recognition and understanding using distant microphonesand omni-directional camera", IEEE Trans. Audio Speech Lang. Process., vol. 20, no. 2, pp. 499-513, Feb. 2012.

8. T. Hori, C. Hori, Y. Minami and A. Nakamura, "Efficient WFST-based one-pass decoding withon-the-fly hypothesis rescoring in extremely large vocabulary continuous speechrecognition", IEEE Trans. Speech Audio Process., vol. 15, no. 4, pp. 1352-1365, May 2007.

9. A. Hyvärinen, J. Karhunen and E. Oja, Independent Component Analysis, USA, NY, New York:Wiley, 2001.

10. C. J. Leggetter and P. C. Woodland, "Maximum likelihood linear regressionfor speaker adaptation of continuous density hidden Markov models", Comput. Speech Lang., vol. 9, no. 2, pp. 171-185, 1995.

11. K. Maekawa, H. Koiso, S. Furui and H. Isahara, "Spontaneous speech corpus of Japanese", Proc. 2nd

Int. Conf. Lang. Resources Eval. (LREC00), pp. 947-952, 2000.

12. K. V. Mardia and I. L. Dryden, "The complex Watson distributionand shape analysis", J. R. Statist. Soc. Ser. B (Statist. Methodol.), vol. 61, no. 4, pp. 913-926, 1999.

13. E. McDermott, S. Watanabe and A. Nakamura, "Discriminative training basedon an integrated view of MPE and MMI in margin and error space", Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP10), pp. 4894-4897, 2010.

14. J. Ming, R. Srinivasan and D. Crookes, "A corpus based approach tospeech enhancement from nonstationary noise", IEEE Trans. Audio Speech Lang. Process., vol. 19, no. 4, pp. 822-836, May 2011.

15. P. J. Moreno, B. Raj and R. M. Stern, "A vector Taylor series approach for environment-independentspeech recognition", Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP96), vol. 2, pp. 733-736, 1996.

16. A. Nádas, D. Nahamoo and M. A. Picheny, "Speech recognition using noise-adaptiveprototypes", IEEE Trans. Acoust. Speech Signal Process., vol. 37, no. 10, pp. 1495-1503, Oct. 1989.

17. T. Nakatani, S. Araki, T. Yoshioka and M. Fujimoto, "Multichannel source separation based onsource location cue with log-spectral shaping by hidden Markov source model", Proc. Interspeech10, pp. 2766-2769, 2010.

18. T. Nakatani, S. Araki, M. Delcroix, T. Yoshioka and M. Fujimoto, "Reduction of highly nonstationary ambientnoise by integrating spectral and locational characteristics of speech andnoise for robust ASR", Proc. Interspeech11, pp. 1785-1788, 2011.

19. T. Nakatani, S. Araki, T. Yoshioka and M. Fujimoto, "Joint unsupervised learning of hidden Markovsource models and source location models for multichannel source separation", Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP11), pp. 237-240, 2011.

20. T. Nakatani, T. Yoshioka, S. Araki, M. Delcroix and M. Fujimoto, "Logmax observation model with MFCC-basedspectral prior for reduction of highly nonstationary ambient noise", Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP12), pp. 4029-4033, 2012.

21. T. Nakatani, M. Souden, S. Araki, T. Yoshioka, T. Hori and A. Ogawa, "Coupling beamforming with spatial and spectral feature basedspectral enhancement and its application to meeting recognition", Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP13), 2013-May.

22. M. H. Radfar, W. Wong, R. M. Dansereau and W.-Y. Chan, "Scaled factorial hidden Markov models: Anew technique for compensating gain differences in model-based single channelspeech separation", Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP10), pp. 1918-1921, 2010.

23. M. G. Rahim and B.-H. Juang, "Signal bias removal by maximumlikelihood estimation for robust telephone speech recognition", IEEE Trans. Speech Audio Process., vol. 4, no. 1, pp. 19-30, Jan. 1996.

24. S. J. Rennie, J. R. Hershey and P. A. Olsen, "Single-channel multitalkerspeech recognition", IEEE Signal Process. Mag., vol. 27, no. 6, pp. 66-80, Nov. 2010.

25. S. T. Roweis, "Factorial models and refilteringfor speech separation and denoising", Proc. Interspeech03, pp. 1009-1012, 2003.

26. H. Sawada, S. Araki and S. Makino, "Underdetermined convolutive blind sourceseparation via frequency bin-wise clustering and permutation alignment", IEEE Trans. Audio Speech Lang. Process., vol. 19, no. 3, pp. 516-527, Mar. 2011.

27. M. L. Seltzer and R. M. Stern, "Subband likelihood-maximizingbeamforming for speech recognition in reverberant environments", IEEE Trans. Audio Speech Lang. Process., vol. 14, no. 6, pp. 2109-2121, Nov. 2006.

28. M. Souden, J. Chen, J. Benesty and S. Affes, "An integrated solution for online multichannel noise trackingand reduction", IEEE Trans. Audio Speech Lang. Process., vol. 19, no. 7, pp. 2159-2169, Sep. 2011.

29. D. H. Tran-Vu and R. Häb-Umbach, "Blind speech separation employingdirectional statistics in an expectation maximization framework", Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP10), pp. 241-244, 2010.

30. E. Weinsterin, M. Feder and A. V. Oppenheim, "Multi-channel signal separation by decorrelation", IEEE Trans. Speech Audio Process., vol. 1, no. 4, pp. 405-413, Oct. 1993.

31. J. Woodruff and D. L. Wang, "Sequential organization ofspeech in reverberant environments by integrating monaural grouping and binaurallocalization", IEEE Trans. Audio Speech Lang. Process., vol. 18, no. 7, pp. 1856-1866, Nov. 2010.

32. O. Yilmaz and S. Rickard, "Blind separation of speechmixture via time-frequency masking", IEEE Trans. Signal Process., vol. 52, no. 7, pp. 1830-1847, Jul. 2004.

33. X. Zhao and Z. Ou, "Closely coupled array processing and model-basedcompensation for microphone array speech recognition", IEEE Trans. Audio Speech Lang. Process., vol. 15, no. 3, pp. 1114-1122, Mar. 2007.