HAIR NOISE ANALYSIS AND MITIGATION FOR SMART GLASSES AUDIO CAPTURES

Subrata Biswas[1,2], Daniel Wong[2], Bashima Islam[1], Sanjeel Parekh[2], Vladimir Tourbabin[2]

[1]Worcester Polytechnic Institute, MA, USA [2]Meta Reality Labs, WA, USA
{sbiswas, bislam}@wpi.edu {ddewong, sanjeel, vtourbabin}@meta.com

ABSTRACT

Head-worn devices such as augmented-reality (AR) and smart glasses introduce a previously overlooked form of audio degradation: hair noise, caused by the wearer's hair brushing against device frames and embedded microphones. To the best of our knowledge, this phenomenon has not been systematically studied. This paper addresses this gap through three contributions. First, we conduct a user study quantifying the perceptual annoyance of hair noise. Second, we introduce the Hair Noise Mitigation (HNM) dataset, the first multi-channel corpus of hair noise collected across diverse real-world conditions. We further characterize its spectral and spatial properties, revealing a non-stationary and directionally dependent nature. Finally, we propose online and offline semi-supervised non-negative matrix factorization (NMF) methods as benchmark mitigation approaches, showing perceptual gains that motivate further research. Together, these contributions establish hair noise as a distinct challenge for wearable audio systems and lay the groundwork for tailored enhancement techniques.

***Index Terms*—** Hair Noise, Smart Glasses, Noise Suppression

1. INTRODUCTION

Augmented-reality (AR) and smart glasses [1] are increasingly enabling users to capture and share immersive audio-visual experiences. Equipped with microphone arrays, these head-worn devices provide spatial audio that enhances realism and presence. However, their wearable form factor introduces a previously overlooked challenge: hair noise, caused by the wearer's hair brushing against device frames and microphones. Such noise arises during everyday actions such as brushing, adjusting, or moving hair, as well as rapid head motions, and can significantly compromise audio quality and user experience. Its severity and characteristics vary across users and devices, motivating a closer examination of the underlying factors.

Hair noise depends on several factors, including hair length and texture, user movement, microphone array design, and device fit. Unlike background noise such as conversations or traffic, hair noise originates from the wearer's own interactions with the device and occurs in close proximity to the microphones, often leading to severe degradations in signal quality. Recordings containing hair noise typically exhibit a mean SNR of $\approx$8.5 dB for the wearer's speech, with subjective evaluations rating them from slightly annoying to clearly annoying; in extreme cases, the SNR can drop to nearly 0 dB, severely diminishing intelligibility and overall quality.

This issue has also been noted in hearing aids, where users often resort to improvised hardware fixes shared in online forums and videos [2, 3]. Yet, to the best of our knowledge, no prior work has systematically studied the acoustic characteristics of hair noise or explored machine-learning-based strategies for its mitigation. We address this gap through the following contributions.

Fig. 1: Ray-Ban Meta microphone positions, channel numbers and porting directions (indicated by green arrows).

Perceptual Study of Hair Noise. We conduct a controlled user study to quantify the impact of hair noise on speech quality. Our results show that hair noise becomes perceptually disruptive below 5 dB SNR, with ratings dropping to "annoying," while annoyance diminishes substantially above 15 dB.

Hair Noise Mitigation (HNM) Dataset and Characterization. We release the first open-source dataset dedicated to hair noise, consisting of 5-channel recordings collected with Ray-Ban Meta (RBM) smart glasses [1]. The dataset captures diverse noise patterns from head movements and hand interactions, enabling spatial and spectral characterization. Our analysis reveals distinct acoustic patterns tied to microphone placement, highlighting the non-stationary and directionally dependent nature of hair noise.

Hair Noise Suppression Benchmark. We establish benchmark suppression results using semi-supervised Non-negative Matrix Factorization (NMF) in both offline and online modes. NMF [4] learns spectral signatures of hair noise to separate it from speech, providing a practical baseline for developing advanced enhancement methods in AR/VR systems and complements existing approaches such as spectral subtraction [5] and deep learning–based enhancement techniques [6, 7, 8].

The remainder of the paper is organized as follows: Section 2, 3, and 4 cover the user study, dataset collection, hair noise analysis, and mitigation method. Section 5 presents results, including evaluations of offline and online NMF, user study findings, and an ablation study for the choice of key hyperparameters.

2. PERCEPTUAL USER STUDY OF HAIR NOISE

To understand how hair noise affects subjective audio quality, we conducted a listening test with 16 participants. Each participant rated 10 randomized samples of own-voice recordings, where clean speech was mixed with hair noise at predefined SNRs ranging from -5 dB to 20 dB. Ratings were given on a 5-point annoyance scale, where 1 indicates "very annoying" and 5 indicates "imperceptible." Figure 2 shows the distribution of ratings across SNR levels. As expected, samples at -5 dB and 0 dB received predominantly low

Work done during Subrata Biswas' internship at Meta Reality Labs.

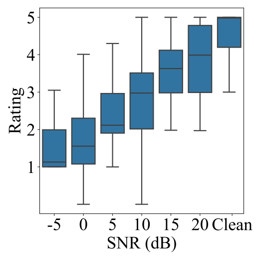
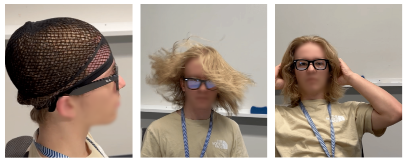


Fig. 2: User perception ratings when `Own-Voice` is obstructed by hair noise.



(a) Clean speech while wearing wig-cap (b) Hair noise due to haking head (c) Hair noise due to playing with hair

Fig. 3: Layered data collection setup to obtain clean speech and hair-noise (hair-playing and head-shaking) recordings.

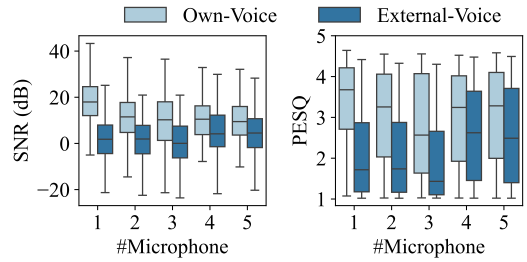


Fig. 4: Effect of hair-noise on the recorded `Own-Voice` and `External-Voice` due to the placement of microphones on RBM.

ratings (median rating of $1-2$), indicating strong perceptual discomfort. Ratings improved progressively with increasing SNR, clustering around "neutral" (3) at 10 dB and exceeding 4 at $15-20$ dB. Clean speech consistently achieved ratings near the maximum value of 5. These results suggest that hair noise becomes perceptually disruptive below 5 dB SNR, while annoyance diminishes substantially at 15 dB or higher.

Importantly, these findings demonstrate that hair noise is not a negligible artifact, but a perceptually valid problem. Even mild interference produces annoyance levels comparable to other well-studied noise sources, and in extreme cases (e.g., 0 dB SNR), hair noise can severely impact signal quality. Since AR and smart glasses are increasingly promoted as devices for always-on voice capture and immersive communication, the presence of such user-generated noise poses a real barrier to achieving high-quality, reliable audio. This motivates the need for systematic datasets, detailed characterization, and the development of robust suppression methods tailored specifically to hair noise.

3. HNM DATASET AND HAIR NOISE ANALYSIS

To enable systematic study of hair noise, we collected the first Hair Noise Mitigation (HNM) dataset, recorded using Ray-Ban Meta (RBM) smart glasses equipped with a five-microphone array [1]. Figure 1 illustrates the microphone placement on the RBM device. Recordings were conducted across varied acoustic environments, ranging from anechoic chambers to rooms with reverberation times (RT60) up to 600 ms, ensuring coverage of both controlled and realistic conditions.

Table 1: Hair Noise Mitigation (HNM) Dataset Summary

	Category	Amount	Duration
Speech	`Own-Voice`	32	$32 \times 60s = 1920s$
	`External-Voice`	34	$34 \times 60s = 2040s$
	Total	66	$3960s$
Noise	Hair-Playing	51	$51 \times 60s = 3060s$
	Head-Shaking	51	$51 \times 60s = 3060s$
	Total	102	$6120s$

Data Collection Protocol. We adopted a *layered collection protocol* to isolate clean speech and hair-generated noise, facilitating systematic analysis and usage. Participants first wore a secure wig cap to suppress hair noise during speech recording as shown in Fig. 3a. While wearing the glasses, they read aloud from a script, producing clean speech samples. In a second phase, the wig cap was removed and participants performed specified movements such as shaking their heads or playing with their hair to introduce natural hair-generated noise, resulting in two distinct types of hair noise captured in the dataset (Figs. 3b and 3c).

In addition to the two types of hair noise, the dataset also includes two categories of voice recordings: `Own-Voice` (the participant's speech) and `External-Voice` (speech from another speaker a few meters away). This design enables analysis of hair noise effects in both near-field and far-field scenarios. In this work, we focus specifically on speech as the signal of interest, leaving the inclusion of other ambient sounds to future work. In total, 17 participants (4 male, 13 female) contributed to the dataset, with recordings sampled at 48 kHz and 32-bit depth. Table 1 summarizes the dataset's overall size and distribution.

Impact of Microphone Position on Speech Quality. Hair noise primarily affects the lower frequency spectrum (< 1 kHz), overlapping with fundamental speech frequencies and degrading intelligibility. Our analysis shows that microphone placement strongly influences susceptibility to hair noise. Temple microphones (channels $\#2$ and $\#3$ in Fig. 1), positioned near the hairline, are most vulnerable to brushing and tangling, leading to lower Signal-to-Noise Ratio (SNR) and Perceptual Evaluation of Speech Quality (PESQ) [9] scores. By contrast, the nose microphone (channel $\#1$), positioned away from the hairline, consistently yielded cleaner recordings. Figure 4 quantifies this effect, showing severe quality drops for temple microphones, particularly in the `External-Voice` condition where hair noise dominates over distant speech.

Spectral and Temporal Characteristics of Hair Noise. Our analysis reveals that hair noise exhibits distinct spectral and temporal signatures. Playing with hair produces persistent, low-frequency noise that continuously overlaps with speech, reducing overall quality. In contrast, head-shaking generates intermittent, pulsating noise aligned with motion dynamics. This bursty nature suggests that suppression algorithms could exploit temporal sparsity to improve effectiveness. Figure 5 illustrates the contrasting spectrograms of clean, noisy, and hair-contaminated signals, highlighting the differences between persistent and intermittent hair noise patterns. These observations confirm that hair noise is not random but exhibits structured patterns that must be considered in mitigation strategies.

In summary, as the first dataset of its kind, HNM offers valuable insights into the characteristics of hair noise and should help direct future work on the problem. The dataset's size makes classical ML approaches such as NMF and fine-tuning pretrained networks preferred mitigation strategies over training a model from scratch.

4. HAIR NOISE SUPPRESSION BENCHMARK

We benchmark hair noise suppression using a semi-supervised Non-negative Matrix Factorization (NMF) framework [10, 11]. NMF is

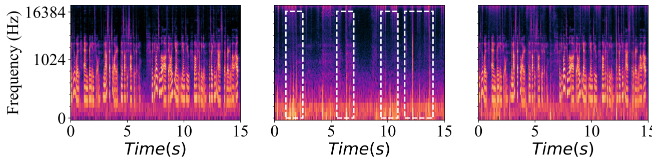


Fig. 5: Clean `Own-Voice` spectrogram (left). Hair noise spectrogram from head-shaking (middle). Dashed boxes mark hair noise instances. Noisy `Own-Voice` spectrogram (right).

chosen because (i) it is a well-established method in audio source enhancement, (ii) its interpretable structure provides insight into the spectral-temporal signatures of hair noise, and (iii) it can be adapted for both offline and online operation, making it relevant for AR devices with real-time constraints. Our proof-of-concept implementation focuses on single-channel enhancement, exploiting the distinct spectral and temporal patterns of hair noise.

Given a noisy speech signal $x(t) = s(t) + n(t)$, where $s(t)$ and $n(t)$ denote the time-domain clean speech and noise signals, we first apply the Short-Time Fourier Transform (STFT) to obtain a time-frequency representation. Let $\mathbf{X}(f,t)$ denote the STFT of $x(t)$, with magnitude spectrogram $\mathbf{V} = |\mathbf{X}| \in \mathbb{R}_{\geq 0}^{F\times T}$, where F is the number of frequency bins and T the number of time frames.

NMF seeks to approximate $\mathbf{V}$ by decomposing it into two low-rank non-negative matrices: a dictionary matrix $\mathbf{W} \in \mathbb{R}_{\geq 0}^{F\times K}$ and an activation matrix $\mathbf{H} \in \mathbb{R}_{\geq 0}^{K\times T}$, where K is the number of basis vectors. This can be expressed as $\mathbf{V} \approx \mathbf{WH}$. Here, $\mathbf{W}$ captures frequency patterns (e.g., speech vs. noise), while $\mathbf{H}$ encodes their temporal activations by describing the time-varying contributions of each frequency pattern. The optimization objective is:

$$\min_{\mathbf{W},\mathbf{H}} D(\mathbf{V}|\mathbf{WH}) \text{ subject to } \mathbf{W} \geq 0, \mathbf{H} \geq 0 \tag{1}$$

Here, $D(.|.)$ is a separable divergence such as Euclidean distance, Kullback-Leibler (KL) divergence, or Itakura-Saito (IS) divergence. $\mathbf{W}$ and $\mathbf{H}$ are estimated with multiplicative updates [4][1].

In our semi-supervised setting, $\mathbf{W}$ is partitioned into speech ($\mathbf{W}_s \in \mathbb{R}_{\geq 0}^{F\times k_s}$) and noise components ($\mathbf{W}_n \in \mathbb{R}_{\geq 0}^{F\times k_n}$). By construction, this leads to a division of $\mathbf{H}$ into signal ($\mathbf{H}_s \in \mathbb{R}_{\geq 0}^{k_s\times T}$) and noise ($\mathbf{H}_n \in \mathbb{R}_{\geq 0}^{k_n\times T}$) temporal activation matrices, where $K = k_s + k_n$. $\mathbf{V} \approx \mathbf{WH}$ can then be re-organized as:

$$\mathbf{V} \approx \begin{bmatrix}\mathbf{W}_s & \mathbf{W}_n\end{bmatrix}\begin{bmatrix}\mathbf{H}_s \\ \mathbf{H}_n\end{bmatrix} \tag{2}$$

In the training phase, only hair-noise recordings are used to learn $\mathbf{W}_n$ [14]. At inference, $\mathbf{W}_n$ is fixed while $\mathbf{W}_s$, $\mathbf{H}_s$, and $\mathbf{H}_n$ are estimated, forcing the model to represent hair noise using previously learned bases. This constraint improves separation by embedding prior knowledge of hair noise.

The enhanced speech $\hat{\mathbf{S}}$ is obtained via mask-based filtering:

$$\hat{\mathbf{S}} = \frac{\mathbf{W}_s\mathbf{H}_s}{\mathbf{WH}} \odot \mathbf{X} \tag{3}$$

where division and $\odot$ denote element-wise operations. Finally, the inverse STFT reconstructs the time-domain enhanced signal.

For online operation, we adapt the sliding-window approach [10], where NMF decomposition is computed using a limited history of recent frames. This adaptation enables low-latency enhancement, making the method applicable to wearable AR devices. Sec. 5.1 details the hyperparameters, including window length and delay.

[1]See [12, 13] for further details.

Table 2: Performance of Online and Offline NMF on `Own-Voice` and `External-Voice` with different cost functions. Best values are denoted in **bold**.

Signal	Method	SI-SDR (dB) ↑		SI-SIR (dB) ↑	
		Offline	Online	Offline	Online
`Own-Voice`	Noisy	10.62		11.04	
	Spectral Subtraction	6.46	-	9.57	-
	NMF Itakura-Saito	11.16	11.10	19.49	17.44
	NMF Kullback-Leibler	**11.48**	**11.29**	**24.08**	**18.78**
`Ext-Voice`	Noisy	2.51		3.03	
	Spectral Subtraction	-0.75	-	0.91	-
	NMF Itakura-Saito	2.43	1.99	11.12	7.07
	NMF Kullback-Leibler	**3.17**	**3.01**	**12.89**	**6.02**

5. EVALUATION

Experimental Setup. We evaluate our semi-supervised NMF-based hair noise suppression method using both offline and online implementations across two divergence measures: Kullback–Leibler (KL) and Itakura–Saito (IS). These cost functions are widely used in audio: KL emphasizes larger values, while IS is scale-invariant [15].

Throughout this study, microphone #3 is used as the primary channel for both voice and noise signals. For all experiments, a 1024-point DFT is applied to compute the STFT. We set the number of speech and noise components to $k_s = 20$ and $k_n = 5$, respectively (see Sec. 5.5 for ablation justification). For the `External-Voice` condition, the external speaker is treated as the target signal, while the wearer engages in hair-related activities such as shaking or playing with hair.

Performance is measured using Scale-Invariant Signal-to-Distortion Ratio (SI-SDR) and Scale-Invariant Signal-to-Interference Ratio (SI-SIR) [16]. SI-SDR captures all distortions, including both interference and algorithmic artifacts, while SI-SIR isolates the impact of the interfering source, i.e., the hair-noise, excluding artifacts introduced by the separation process.

For evaluation, hair noise samples from each participant are split 80/20 into training and test sets. The training subset is used to learn $\mathbf{W}_n$, and noisy test mixtures are generated by mixing held-out noise with corresponding clean speech recordings.

5.1. Quantitative Results

This section presents the quantitative results for both the offline and online benchmark hair noise suppression algorithms.

Offline NMF Results. Table 2 reports SI-SDR and SI-SIR for offline NMF, comparing noisy and enhanced signals under both cost functions. KL divergence consistently provides the best performance across both conditions and metrics. IS divergence improves over noisy baselines in the `Own-Voice` scenario but falls short in SI-SDR for `External-Voice`. Overall, NMF effectively suppresses hair noise (as reflected in SI-SIR improvements) but may introduce artifacts that reduce overall quality. Analysis by noisy input SI-SDR reveals an interesting trend: NMF provides larger gains for low-SNR scenarios ($\leq$ 5 dB) than for high-SNR cases, as shown in Fig. 6. In comparison, spectral subtraction performs poorly due to non-stationary nature of hair noise.

Online NMF Results. We adapt NMF to an online mode using a sliding-window which contains recent and past signal spectra. The window length is determined by the hyperparameter `delay`, which specifies how many past frames are retained. At each new frame, the window shifts by one frame. We set `delay` = 8, corresponding to 8 past frames (32 ms each), and reduce iterations to 16 for computational efficiency (see Sec. 5.5 for detailed ablation study).

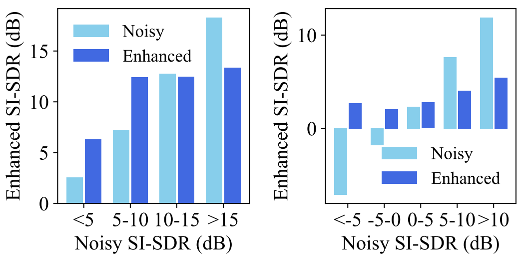


Fig. 6: Signal quality improvement for different noise levels. (Left) `Own-Voice` and (Right) `External-Voice`.

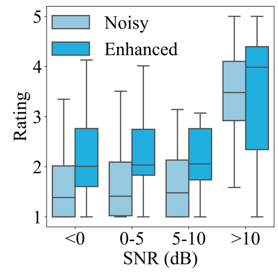


Fig. 7: Perceptual annoyance results for `Own-Voice`. Ratings 1 & 5 correspond to *"very-annoying"* & *"imperceptible."*, respectively.

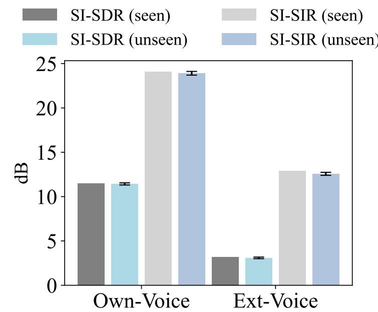
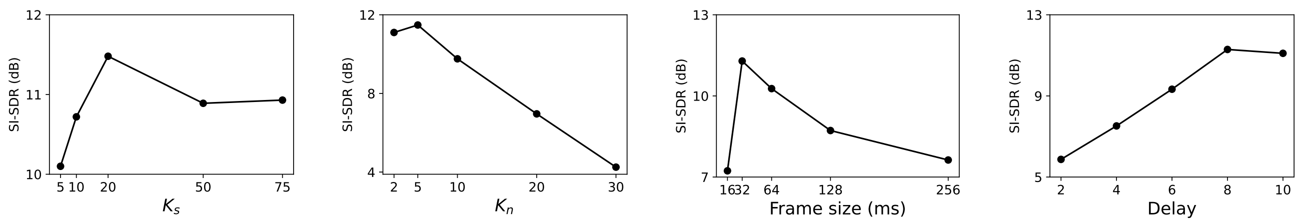


Fig. 8: Generalization of offline NMF with KL divergence to unseen users.



(a) Offline: k_s vs SI-SDR. (b) Offline: k_n vs SI-SDR (c) Online: Frame len vs SI-SDR (d) Online: Delay vs SI-SDR

Fig. 9: Ablation studies on hyper-parameter choices for both offline and online NMF.

Performance trends mirror those of the offline case: KL remains the best divergence measure. As expected, online NMF performs slightly worse than offline due to shorter context. However, the gap is small, for example, SI-SDR in the (KL, `Own-Voice`) condition decreases by only 0.19 dB. This suggests online NMF is feasible for device-constrained scenarios with near-offline performance.

5.2. Subjective Evaluation

We conducted a listening test where 11 participants rated 10 randomly selected noisy and enhanced signals on a $1-5$ annoyance scale. Fig. 7 shows the distribution of rating improvements across SNR ranges. Enhanced samples consistently received higher ratings, with most scores ranging from "neutral" to "imperceptible" $(3-5)$. In contrast, noisy samples were rated "annoying" or "neutral" $(2-3)$. Improvements were particularly pronounced in low-SNR conditions ($\leq$ 5 dB), confirming that NMF enhancement improves perceived listening comfort.

5.3. Generalization to Unseen Users

We tested offline NMF with KL divergence using leave-one-out cross-validation across participants. As shown in Fig. 8, the method generalizes well: performance for unseen users is consistent with Table 2. Low standard deviations in both SI-SDR and SI-SIR demonstrate robustness across participants for both `Own-Voice` and `External-Voice` conditions.

5.4. Computational Complexity

The offline NMF model contains ≈15K parameters, requires 58 KB memory, and has complexity $\mathcal{O}(F\times T\times K\times N)$, corresponding to ≈ 241 MMAC/s at $N = 200$ iterations. Online NMF contains ≈13K parameters, requires 50 KB memory, and has complexity $\mathcal{O}(F \times T_1 \times K \times N \times T_2)$, corresponding to ≈ 59 MMAC/s with $N = 16$, $T_1 = 9$ frames per window, and $T_2 = 32$ windows.

5.5. Ablation Studies

We evaluate the impact of key parameters on the performance of offline and online NMF methods, aiming to optimize SI-SDR while balancing the algorithm's computational complexity.

Effect of k_s in Offline NMF. Varying k_s from 5 to 75 revealed optimal SI-SDR at $k_s = 20$ (Fig. 9a); beyond this, improvements saturated. With $k_s = 20$, varying k_n from 2 to 30 yielded the best performance at $k_n = 5$ (Fig. 9b), with no further gains for larger values.

Effect of Frame Size in Online NMF:. Frame sizes between 16 ms and 256 ms showed optimal performance at 32 ms (Fig.9c); smaller frames lacked context while larger frames oversmoothed. With frame size fixed at 32 ms, varying `delay` from 2 to 10 revealed best performance at 8 frames (Fig.9d). Shorter delays reduced temporal modeling, while longer delays added latency without benefit.

6. CONCLUSION

This paper introduces hair noise as a previously overlooked source of audio degradation in smart glasses. We began with a controlled user study that established its perceptual impact, showing that hair noise becomes disruptive below ≈ 5 dB SNR. To support systematic research, we released HNM, the first multi-channel dataset capturing hair noise across varied real-world conditions, and analyzed its spectral, temporal, and spatial properties, including sensitivity to microphone placement. Building on these insights, we developed a semi-supervised NMF-based suppression method in both offline and online modes, demonstrating measurable improvements in objective metrics and subjective ratings. Together, these contributions establish hair noise as a valid challenge for wearable audio systems and provide a foundation for future development of tailored enhancement techniques.

7. REFERENCES

[1] "Ray-Ban Meta," https://www.meta.com/ai-glasses/, Accessed: 2025-08-04.

[2] "Noises that interfere with hearing aids: A guide for new wearers," https://www.earhealth.co.nz/noises-that-interfere-with-hearing-aids/, Accessed: 2025-09-10.

[3] "Hearing aid rustling/scratching sound," https://www.youtube.com/watch?v=nQvypE4osrw, Accessed: 2025-09-10.

[4] Daniel Lee and H Sebastian Seung, "Algorithms for non-negative matrix factorization," *Advances in Neural Information Processing Systems*, vol. 13, 2000.

[5] Michael Berouti, Richard Schwartz, and John Makhoul, "Enhancement of speech corrupted by acoustic noise," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 1979, vol. 4, pp. 208–211.

[6] Po-Sen Huang, Minje Kim, Mark Hasegawa-Johnson, and Paris Smaragdis, "Deep learning for monaural speech separation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 1562–1566.

[7] Anurag Kumar and Dinei Florencio, "Speech enhancement in multiple-noise conditions using deep neural networks," in *Interspeech 2016*, 2016, pp. 3738–3742.

[8] Hendrik Purwins, Bo Li, Tuomas Virtanen, Jan Schlüter, Shuo-Yiin Chang, and Tara Sainath, "Deep learning for audio signal processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 2, pp. 206–219, 2019.

[9] A.W. Rix, J.G. Beerends, M.P. Hollier, and A.P. Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP).*, 2001, vol. 2, pp. 749–752 vol.2.

[10] Cyril Joder, Felix Weninger, Florian Eyben, David Virette, and Björn Schuller, "Real-time speech separation by semi-supervised nonnegative matrix factorization," in *10th International Conference on Latent Variable Analysis and Signal Separation, LVA/ICA 2012, Tel Aviv, Israel, March 12-15, 2012*. Springer, 2012, pp. 322–329.

[11] Paris Smaragdis, Bhiksha Raj, and Madhusudana Shashanka, "Supervised and semi-supervised separation of sounds from single-channel mixtures," in *International Conference on Independent Component Analysis and Signal Separation*. Springer, 2007, pp. 414–421.

[12] Cédric Févotte, Emmanuel Vincent, and Alexey Ozerov, "Single-channel audio source separation with NMF: divergences, constraints and algorithms," *Audio Source Separation*, pp. 1–24, 2018.

[13] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, and Shun-ichi Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*, John Wiley & Sons, 2009.

[14] Juan José Burred, "Detailed derivation of multiplicative update rules for nmf," *Paris, France*, 2014.

[15] Cédric Févotte, Nancy Bertin, and Jean-Louis Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis," *Neural computation*, vol. 21, no. 3, pp. 793–830, 2009.

[16] Jonathan Le Roux, Scott Wisdom, Hakan Erdogan, and John R Hershey, "SDR–half-baked or well done?," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 626–630.