

# New Technologies of Directional Microphones for Hearing Aids

Xubao Zhang<sup>1,2</sup>

<sup>1</sup>Research & Development, Sonova Unitron, Kitchener, Canada

<sup>2</sup>Electronic Engineering Department, Xi'an Electronic Science and Technology University, Xi'an, China

## Email address:

xbzwdl@yahoo.com

## To cite this article:

Xubao Zhang. New Technologies of Directional Microphones for Hearing Aids. *Journal of Electrical and Electronic Engineering*. Vol. 8, No. 3, 2020, pp. 81-91. doi: 10.11648/j.jeeec.20200803.12

**Received:** May 25, 2020; **Accepted:** June 11, 2020; **Published:** June 23, 2020

---

**Abstract:** This paper describes new technologies of directional microphones for the practical hearing aids, referring to a front-delay direction microphone (DM), narrow beam DM, and minimum variance distortionless response (MVDR) beamformer. Each of the DM technologies was researched against weaknesses of those existing DMs, such as imperfection in low level noise, short suppression to adjacent interference, and failing to simultaneously perceive multiple target voices. In order to eliminate them, the conventional DM architectures have been innovated: the front-delay DM exchanged the elements' positions; the narrow beam DM employed binaural DMs to composite a relatively narrow lobe; the MVDR beamformer combined two types of processing in spatial and frequency domains; and the novel technologies are state-of-the-art beamformers for hearing aids. Based on some references related to the DM technologies and operation principles of the latest beamformers, we further researched the DM technologies, first proposed the implementing architectures, derived new gain equations of the relevant polar plots, accomplished the extensive experiments, and evaluated advantages and disadvantages of the DMs by the obtained evidences; then we confirmed that the new technologies could reach their expected goals. Meanwhile, we used the latest simulating software, Simulink of MatLab R2018b and audio edition software, SoundBooth, in our Lab computers.

**Keywords:** Directional Microphone, Beamformer, Minimum Variance Distortionless Response, Hearing Aid

---

## 1. Introduction

In the noise reduction of hearing aids, a directional microphone (DM) provided greater benefits than other signal processing strategies, so it has been favored by the users and manufactures [1]. Reference [2] did not think that a single DM algorithm could perform with excellence in a changing environment, so he designed a tri-mode DM system with voice-priority. This DM system operated based on the real-time decisions (parallel processing) on environmental sorts, and implemented a quasi-adaptive beamformer. Reference [3] recognized that the traditional DM had lower sensitivity in the low-frequency region than an Omni microphone (mic), thus, they used a processing strategy of Frequency Band Split for their DM system. When the center frequency of some band was lower than the band-split-frequency, the system's Omni mic was activated; otherwise, the traditional DM was activated. Their test results illustrated

that this strategy enhanced the signal-to-noise ratio (SNR), and preserved the voice quality. Reference [4] analyzed and illustrated the existence of spectrum distortion of the traditional DM with human voices by experiments. He designed frequency response balanced DMs, composed of a bandpass filter bank, multipliers, and an adder following a traditional DM. When the multipliers' gains met the requirement of flat frequency response, this balanced DM satisfactorily removed the distortion. Sonova hearing aid DMs featured the balanced DM [5]. The three technologies employed the Omni mic, traditional DMs, and multiband magnifiers etc., and combined them in different ways so as to solve significant low gain in low, mid frequency regions and spectrum distortion. These improved DMs formed the conventional DMs.

Against limitations of performances of the conventional DMs used in practice, hearing researchers and developers have also pursued innovative works on architectures of

traditional DMs. The first one is the front-delay DM, which is a variation of the traditional DM circuit. In soft and low level noises, its speech enhancement and output SNR were better than the Omni mic and the traditional DM. The front-delay DM also was the Soft-Level DM, which was developed by Siemens hearing manufacturer [6]. When power of a situation noise was lower than a certain sound pressure level (SPL), the hearing aid was switched to the Soft-Level DM mode so as to achieve better speech recognition. The second one is the narrow beam DM, which weightily sums the outputs of binaural DMs to achieve a composite beamformer of the end-fire array plus boardside array. With appropriate weights, a relatively narrow beam lobe is formed. Thus, it can more effectively suppress interfering sounds nearby the target source. Phonak and Siemens manufacturers applied this technology in their products [7, 8]. The third one is application of minimum variance distortionless response (MVDR) beamformer [9] to the hearing aid. A main lobe beamformer introduces multi-azimuth target voices and interfering noises, and the beamformer's output subtracts the noise estimator output so as to extract the desired voices. By adjusting the amplitude and delay time of noise estimator output in the frequency bands, the noise reduction of the MVDR beamformer can be maximized, and the multi-azimuth target signals can be extracted well. Additionally, a following frequency domain processing, silent-interval spectral subtraction, removes the remaining noise well. Oticon manufacturer researched and developed Open Sound Navigator (OSN), which was the first creative application of the MVDR beamformer [10]. The existing references have briefly introduced the operation principles, performance evaluations, and field trial verification of their DMs. This paper researched these new DM technologies extensively, including presentations of the detailed principles, demonstration of processing architectures, especially performance verifications with the experimental data and curves obtained by the latest Simulink of MatLab R2018b, and audio edition software in our Lab computers. Now, we assembled many of these research findings for this publication, so as to share with relevant audiologists and hearing aid researchers.

In the end, the appendix discusses the algorithm of a novel signal-to-noise ratio (SNR) estimation, and demonstrates simulation results. The estimation can be used for the noise remover of the MVDR beamformer.

## 2. Common DMs and Sensitivity Gain

We need to briefly illustrate some fundamental concepts related to common DMs before describing the new DM technologies. Figure 1 shows the basic architecture of a cardioid (or hyper-cardioid) DM, composed of two Omni mics, a subtracter, and a time-delay unit. In Figure 1, the solid arrows represent  $0^\circ$  incidence from the front; the dashed arrows represent nonzero incidence ( $\theta$ ). The time parameter  $\tau$  of delay unit determines whether it is a cardioid or hyper-cardioid DM. The cardioid DM is a theoretical concept.

During early implementation of DM hearing aids, Figure 1 operation was the core of a traditional DM. During their late implementation, it was always combined with multiband processing or an Omni mic to complete necessary gain control in the frequency domain, then the traditional DMs transited to conventional DMs.

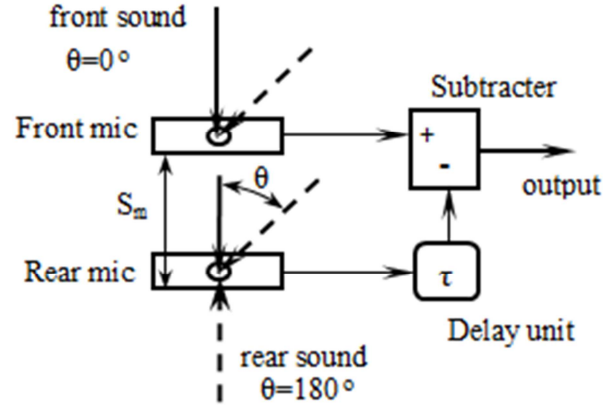
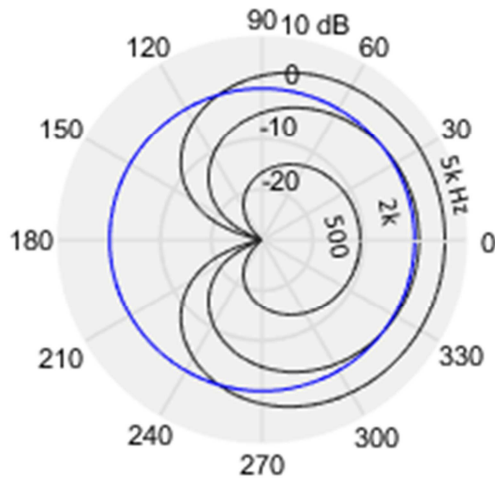


Figure 1. Basic architecture of a cardioid DM.

As is known to all, an Omni mic and a DM both have a basic specification: Sensitivity curve, its unit is dB re 1V/Pa, Pa is a sound pressure unit [11]. For a DM, as beamformer, our first concern was its polar plot. In order to conveniently research the directionality behaviors, we defined DM sensitivity/Omni mic sensitivity as DM sensitivity gain (S-gain), and its unit was dB only. We also derived an equivalent calculating equation for the S-gain: DM output/ Omni mic output, which makes the calculation of DM polar plots simple and rigorous. Assuming that in a sound field, there exists a pure tone of frequency  $f$ , and given delay time  $\Delta$  of the inlet spacing  $S_m$  of the front and rear mics, then  $\Delta = S_m/V_s$ ,  $V_s$  is sound propagation velocity in air, and delay time of the output signals of the two mics  $\delta(\theta) = \Delta \cos(\theta)$ . The S-gain of Figure 1 DM can be derived as:

$$G_{dm}(\theta, f) = 2\sin\{\pi f [\tau + \Delta \cos(\theta)]\} \quad (1)$$

which is related to the pure tone frequency, the inlet spacing of two mics, and the sound incident angle. When  $\tau$  and  $\Delta$  are equal, the obtained DM polar plot is cardioid type. Figure 2 shows S-gain polar plots of a cardioid DM (black) and an Omni mic (blue), the DM parameters:  $\tau = \Delta = 0.04662$  ms,  $S_m = 16$  mm. We can observe that the DM plots are cardioid. The outer black plot results from a tone of 5 kHz, with gain 6 dB at  $0^\circ$ ; the inner black plot, from a tone of 500 Hz, with gain -10.7 dB at  $0^\circ$ . The polar plot of Omni mic is round, with gain 0 dB at  $0^\circ \sim 360^\circ$ , independent of frequencies. From Figure 2, we can obtain that (1) the S-gain plots of the cardioid DM all have a deep notch at  $180^\circ$ , independent of frequencies. This can achieve a strong suppression to the back noises. (2) At  $0^\circ$ , with a frequency  $\geq 1.78$  kHz, the DM S-gain is higher than the Omni mic S-gain; and vice versa. (3) The DM S-gain at  $0^\circ$  is much different as the frequency varies, such a gain response will cause significant spectrum distortion when a speech is going through the DM [4].

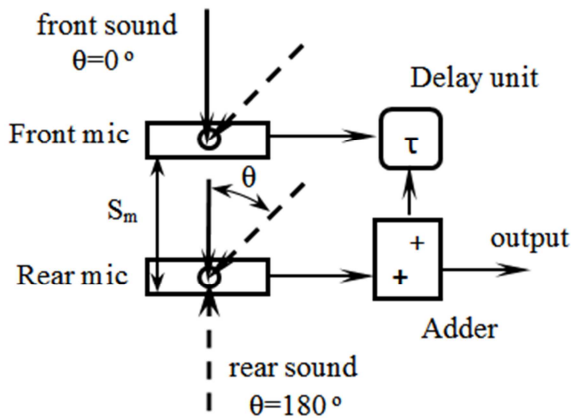
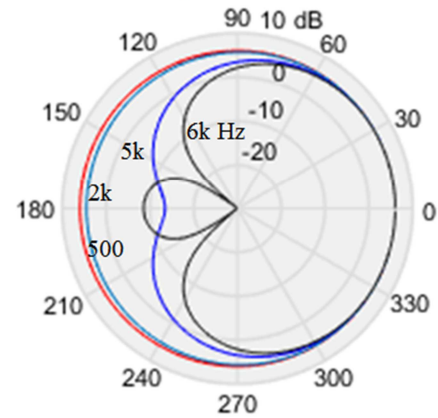
**Polar plots of cardioid DM and Omni mic****Figure 2.** S-gain polar plots of cardioid DM and Omni mic.

### 3. Front-Delay DM

Circuit architecture of the front-delay DM is a variation of Figure 1: the subtracter exchanges the positions with the delay unit, and the subtracter is replaced with an adder, as shown in Figure 3. Assuming that these parameters are the same as those in Figure 1, the S-gain equation of the front-delay DM can be derived as:

$$G_{\text{fdm}}(\theta, f) = 2\cos\{\pi f [\Delta\cos(\theta) - \tau]\} \quad (2)$$

which is related to the pure tone frequency, the inlet spacing of two mics, and the incident angle. Figure 4 shows polar plots of the front-delay DM,  $\tau = \Delta = 0.04662$  ms,  $S_m = 16$  mm. The four plots result from four tones of 6, 5, 2, and 0.5 kHz, respectively. We can observe that the four plots have the same gain 6 dB at  $0^\circ$  incidence, independent of the frequencies; their notches at  $180^\circ$  have different depths, depending on the frequencies: at 5 kHz, the notch is -13.5 dB; at 500 Hz, the notch almost disappears. Thus, the front-delay DM is not effective in suppression to the back noise, but it does not cause spectrum distortion of voices, so does not need to have the frequency response balanced.

**Figure 3.** Basic architecture of a front-delay DM.**Polar plots of front-delay DM in pure tones****Figure 4.** S-gain polar plots of a front-delay DM.

Since the results from pure tones could not illustrate performance in real-world voices, we conducted tests with human voices to clarify. The large-sample data of human voices and real-world noises could easily be acquired from online wave files, whose time series last several seconds or longer. For high confidence, we intercepted an eleven words' voice from [12], "Hi, one of the available high quality texts to speech voices." This voice lasted about 3.8 s, 168,769 samplings, wordlength 16 bits. In the tests below, we used this voice level as the criterion of SPL 60 dB for calibrating. The traffic noise was acquired from a wave file [13], and the white noise as equipment noise from test software [14]; both lasted 4 s, 176,398 samplings. By adjusting RMS of the two noises, we calibrated the noises SPL 50 dB. The voice came at  $0^\circ$  incidence. Considering the shadow of head effect, we selected  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ , and  $180^\circ$  incidences only to simulate the surrounding noise environment. As for the wave file reading, noise SPL calibrating, and simulating procedures, refer to reference [15]. Table 1 lists the outputs and SNRs of the Omni mic, conventional DM, and front-delay DM in the low level noise. When comparing the DM performances of the new technologies in this paper's tests, we selected a frequency response balanced DM as the conventional DM. From this table, it can be obtained that (1) voice enhancement of Omni mic was 0.05 dB (from filter bank), the conventional DM, 1.89 dB, and the front-delay DM, 6.06 dB. (2) In the traffic noise, the conventional DM output (RMS) was lower, the front-delay DM, higher, and the Omni mic, mid; when plus the voice, the conventional DM got SNR 13.3 dB, the front-delay DM, 11.4 dB, and the Omni mic, 11.2 dB. (3) In the equipment noise, the conventional DM output was mid, the front-delay DM, higher, and the Omni mic, lower; when plus the voice, the conventional DM got SNR 14.6 dB, the front-delay DM, 17.1 dB, and the Omni mic, 15.4 dB. The data indicate that in the low (or soft) level noise environment, the two DMs and Omni mic all got enough SNR to understand the noisy voice. The front-delay DM got better SNR than the Omni mic in the both noises, but the conventional DM got better SNR than the Omni mic only in the traffic noise. Summarizing those plots in Figures 2, 4 and these data in Table 1, we can conclude that the

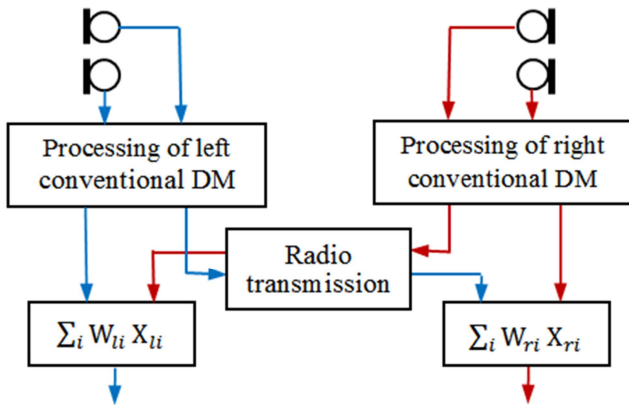
front-delay DM and Omni mic all performed the same preservation of speech naturalness, but the front-delay DM got the better SNR than the Omni mic in the low/soft level noises. Therefore, the front-delay DM can undoubtedly replace the Omni mic. The front-delay DM is also a type soft-level DM.

**Table 1.** Outputs and SNRs of Omni mic, conventional DM, and front-delay DM in low level noises.

Sorts of sounds	Omni mic	Conventional DM	Front-delay DM
Voice enhancement (dB)	0.05	1.89	6.06
Traffic noise (RMS)	0.0135	0.013	0.0267
SNRs (dB)	11.2	13.3	11.4
Equipment noise (RMS)	0.00823	0.0112	0.0136
SNRs (dB)	15.4	14.6	17.1

## 4. Narrow Beam DM

In some complex listening environment, target voice source and masked noise source are not separated far away in azimuth, e.g., their azimuth difference is  $60^\circ$  or less. The conventional DM cannot get the output SNR enough. This is because its beam lobe has a large width of about  $\pm 90^\circ$ . To conquer this shortcoming, reference [7] developed the narrow beam DM, which operated the outputs of bilateral conventional DMs by a weighted sum, as shown in Figure 5. The priori optimized weight coefficients  $\{W_{li}\}$  and  $\{W_{ri}\}$  enabled the composite beamform to have a narrow main lobe and flat frequency response. In the binaural DM design, the conventional DM of each side is composed of two (front and rear) mics in longitudinal array, i.e., so-called end-fire linear array, and its beamform is left and right symmetric. On the other hand, the conventional DMs in the sides constitute a boardside linear array, and their beamform is front and back symmetric. Transverse spacing of the two human ears is relatively broad,



**Figure 5.** Block operation diagram of a narrow beam DM.

e.g., spacing between the ears  $S_e=16$  cm. When the weights  $\{W_{li}\}$  and  $\{W_{ri}\}$  are 1s, the S-gain of the boardside binaural DMs can be derived as:

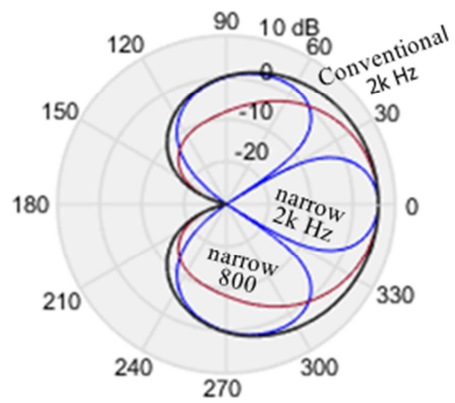
$$G_{bsbf}(\theta, f) = 2\cos[\pi f \sigma \sin(\theta)] \quad (3)$$

where  $\sigma$  is delay time caused by the spacing between two ears. The directionality and frequency response of the boardside

array (3) are much different from those of a conventional DM. S-gain of the narrow beam DM,  $G_{nbf}$ , is product of the two separated DMs' S-gains,  $G_{bsbf}$  and  $G_{efbf}$ .  $G_{efbf}$  is the end-fire array's S-gain. The beam patterns of the boardside array are not as simple as a linear end-fire array, but its frequency response is not sloped, i.e., no gain declines at low, mid frequencies. In pure tone field, (3) can form multiple beams when the tone frequency is high, but a single beam when the frequency is low. Thus, lobe width of the composite beam narrows at low, mid frequencies. Additionally, in implementation, this technology needs a two-way radio transmitter of super low power, which transmits the data of each lateral DM output to the contralateral DM to calculate the weighted sum, i.e., the second directional operation.

Figure 6 shows polar plots of a narrow beam DM and conventional DM in pure tone field,  $S_e=16$  cm,  $S_m=16$  mm, and  $\tau=\Delta=0.04662$  ms. We can observe that a quite narrow lobe (red) results from a tone of 800 Hz; and an extremely narrow lobe (blue) results from a tone of 2 kHz, but two high, wide side lobes accompany it. For a speech, its spectrum contents cover a wide audio range, but the spectrum energy mostly distributes within the low and mid frequency regions. So, the overall lobe of the narrow beam DM would surely be narrower than the main lobe of a conventional DM? This issue needs to be verified by measuring in a real-world voice field. When we verified, the voice was acquired the same as done in Table 1, but the voice was of SPL 60 dB, coming from eight azimuths, the azimuth interval  $45^\circ$ .

**Polar plots of narrow beam DM and conventional DM in pure tones**



**Figure 6.** Polar plots of narrow beam DM and conventional DM in pure tones.

After measuring, the output data of the narrow beam DM and conventional DM were interpolated, then 16 azimuths' data were obtained for each DM. Figure 7 shows the polar plots of the two DMs in the voice field. We can observe that with a cut-off of 3 dB gain drop, the lobe width of the conventional DM was  $\pm 90^\circ$  while the lobe width of the narrow beam DM was  $\pm 60^\circ$ , much narrower than that of the conventional DM. Table 2 lists output SNRs of the conventional DM and narrow beam DM in the same voice and traffic noise as those in Table 1, except that the two



sounds had competing level, SPL 60 dB. We can observe that with azimuths of the traffic noise, 45°, 90°, and 135°, output SNRs of the conventional DM were 0.57, 3.98, and 9.92 dB, respectively, while output SNRs of the narrow beam DM were 4.13, 8.63, and 13.5 dB, respectively. These test data

indicated that the narrow beam DM improved SNR by 3.5~4.5 dB over the conventional DM in the field of common voice plus traffic noise. Additionally, the ears' spacing is larger, the SNR improves more.

Table 2. SNRs of narrow beam and conventional DMs' outputs.

Directional mics	Azimuths between target voice and traffic noise *		
	45°	90°	135°
Conventional DM SNRs (dB)	0.57	3.98	9.92
Narrow beam DM SNRs (dB)	4.13	8.63	13.5

\*Traffic noise and target voice both were of SPL 60dB, Se=16cm.

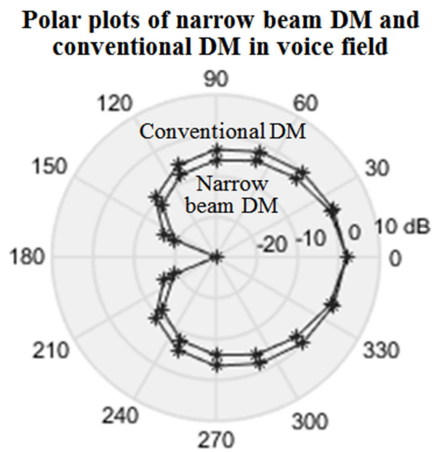


Figure 7. Polar plots of narrow beam DM and conventional DM in voice field.

## 5. Minimum Variance Distortionless Response Beamformer

In recent years, the minimum variance distortionless response (MVDR) beamformer has come out in other electronic industry, it is effective to detect desired multi-point signals in spatial interference [9]. In some listening situations, there exist multiple voices near the target voice while the aid wearer likes to listen to a few of them. These desired voices usually locate in the front or sides of a listener. The interfering noise may be a point or surrounding source, which usually locate in the back or sides. The voices and noise both always have wide spectra, so their spectra often overlap in the audio region. Under such a priori conditions of the listener, it is most suitable to use the MVDR principle for the multi-voice DM design.

### 5.1. Speech Signal Extractor

To introduce voices in such a site, e.g., a meeting hall, a simple Omni mic is assigned for the main beamformer, which receives both the voices and noise within 0°~360° range. The block matrix, which blocks the voice signal, is implemented with a reverse cardioid DM (RDM); as a noise estimator, then its output should be the noise only. The main beamformer output subtracts the noise estimator output to

extract the speech signal. The architecture of the MVDR speech extraction is shown in Figure 8. The main beamformer is composed of the Omni mic, followed by a frequency band. The noise estimator is composed of the middle and bottom blocks in left of Figure 8, followed by a frequency band; in fact, it was the same as in Figure 1, except that the subtractor exchanges the positions with the delay unit. The phase factor  $Z^t$  and block factor  $\alpha$  are adjustable, depending on the band with the factors and the speech/noise detector's decision. The default  $t$  and  $\alpha$  are 0 and 0.5, respectively. The external delay is the delay time between the outputs of the rear and front mics, and is not a real element, denoted with a dashed rectangle. Another delay unit above is an internal element, the same as that in Figure 1. When  $\tau=\Delta$ , the noise estimator is a reverse cardioid DM, and its S-gain equation can be derived as:

$$G_{rdm}(\theta, f) = 2\sin\{\pi f \Delta [\cos(\theta) - 1]\} \quad (4)$$

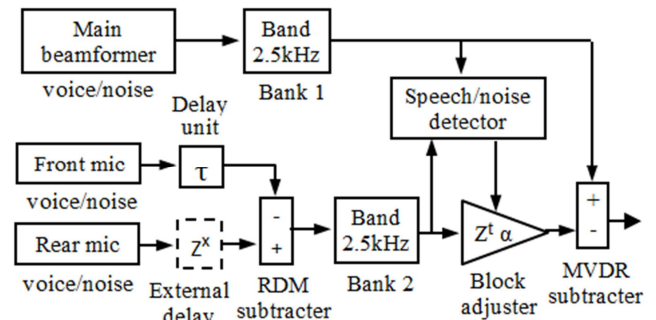


Figure 8. Speech signal extractor of a MVDR beamformer in bands 2.5 kHz.

In the experiment, the optimized noise estimation required independently adjusting  $\alpha$  and  $Z^t$  in most frequency bands, so as to well match the estimator output noise with the output noise of the main beamformer. Under the same test conditions as those of Figure 2, we calculated polar plots of the noise estimator in terms of (4), as shown in Figure 9. The three solid plots resulted from three tones of frequencies 5, 2, and 0.5 kHz, respectively; the dashed polar plot resulted from the tone 500 Hz with the optimized parameters  $\alpha$  and  $t$ . Obviously, these solid plots were 180° turnover of those plots of the traditional DM in Figure 2. By means of analysis, the noise estimator could estimate the back noise at 180° most

sensitively, and estimate the noise from other azimuth, e.g., 90°, less sensitively.

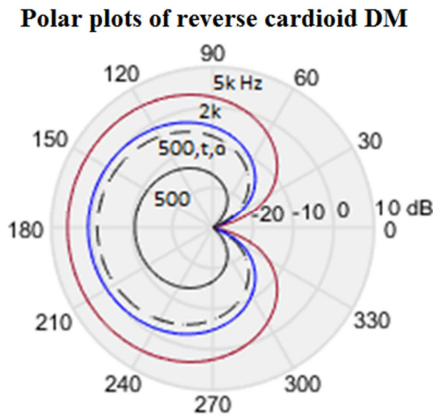


Figure 9. Polar plots of a reverse cardioid DM.

When the speech/noise detector determined NO SPEECH, both  $\alpha$  and  $Z^i$  values were adjusted to be larger in mid and low frequency bands; when determined SPEECH,  $\alpha$  was adjusted to be less so as to preserve the speeches of the main beamformer output; when determined SPEECH PLUS NOISE,  $\alpha$  and  $Z^i$  were optimized, depending on the speech and noise characteristics. In each band, the speech/noise detector always detected the existence of the speech and noise not only from the main beamformer output but also from the RDM estimator output. The speech/noise detector would use the principle of modulation detection with high efficiency. For the details, refer to reference [16]. The MVDR extractor featured: (1) When the main beamformer output subtracted the noise estimator output, the speech signal from the front was not affected at all, so the speech distortion was not caused. (2) When the voice came from the sides, the speech detector could control  $\alpha$  to drop to preserve the side voice outputs in each band. (3) In the environment of noise plus voices, the block factor  $\alpha$  and delay factor  $Z^i$  were adjusted up automatically, then the output noise went down accordingly, and enabled the multi-azimuth voices to be clear and balanced. When fitting a hearing aid, the audiologist can set ranges of  $Z^i$  and  $\alpha$  in each frequency band separately to meet the situation of an individual wearer.

After the outputs of all the frequency bands were summed, full spectrum contents of the MVDR speech output could be integrated. Table 3 lists outputs and SNRs of the Omni mic, conventional DM, and MVDR beamformer in a traffic noise plus two voices. These noise and voices were selected the same as those in Table 2. The first voice came along azimuth 0°, and the second voice, along 90° in the separated tests. The noise resulted from surrounding traffic along azimuths 45°, 90°, 135°, and 180° in both tests. This ignored the opposite noise in the other four azimuths, which were shaded by the wearer's head. In the tests, the range of delay factor  $t$  was 0~14 sampling periods, and  $\alpha$ , within 0.5~1.4, depending on the frequency bands. From Table 3, we can observe that (1) when only voice existed, speech enhancement of the Omni

mic was 0.05 dB (gain of the multiband filter bank); the conventional DM, 1.89 dB (voice 0°) and -3.43 dB (90°); the MVDR beamformer, 0.05 dB (voice 0°) and -0.09 dB (90°). (2) In the traffic noise environment, output (RMS) of the Omni mic was higher, the conventional DM, lower, and the MVDR beamformer, mid; when plus the voices, the Omni mic performed SNR 1.14 dB (voice 0°, 90°), the conventional DM, 6.54 dB (0°) and 1.23 dB (90°), and the MVDR beamformer, SNR 4.15 dB (0°) and 4.1 dB (90°). These data indicated that the MVDR beamformer got better or balanced SNRs than the Omni mic and conventional DM in the open voice situation, i.e., the range of hearable voice azimuths was extended.

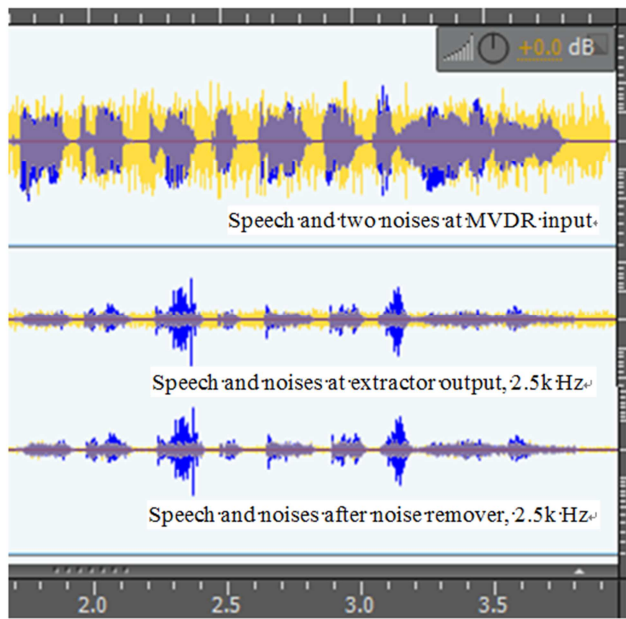
Table 3. Outputs and SNRs of Omni mic, conventional DM, and MVDR beamformer in 60dB noise field.

Sorts of sound	Omni mic	Conventional DM	MVDR beamformer
Voice only (RMS)	0.0487 (0°, 90°)	0.0601 (0°) 0.0326 (90°)	0.0487 (0°) 0.0479 (90°)
Enhancement (dB)	0.05 (0°, 90°)	1.89 (0°) -3.43 (90°)	0.05 (0°) -0.09 (90°)
Traffic noise	0.0427	0.0283	0.0302
SNRs (dB)	1.14 (0°, 90°)	6.54 (0°) 1.23 (90°)	4.15 (0°) 4.1 (90°)

## 5.2. Remaining Noise Remover

In fact, when a surrounding noise had a power competing with a target voice, SNR of the MVDR beamformer output was not satisfactory to well understand the speech, as those SNR data listed in Table 3. Thus, the next processing of the MVDR beamformer required removing the remaining output noise. Usually, when a voice and noise come up, we can observe that the noise signal is moving at any time, but the voice waveform is not that, e.g., in the silent intervals between two adjacent phones or the periods after word/sentence stops, the voices always reveal weak level speech or nothing. Such a distinction is more obvious when viewing the voice and noise in a single frequency band, particularly in higher frequency bands. Then, it may be more effective to remove the interval noise in each split band. Analytically, it is no help to use the traditional spectral subtraction by adjusting a band gain [17]. Based on SNR tested in the silent interval of a voice waveform, a modified spectral subtraction was proposed; instead of reducing the band gain, it reduces the gain of the speech and noise only around the silent intervals. We called such processing the silent-interval spectral subtraction. Additionally, we could not reduce the gains rapidly even when the noise only existed, so as to listen to natural sounds and to preserve the possible voice cues. This needed to control the gain in each silent interval according to the SNRs. Usually, the gain reduction can be divided as a few levels, and the maximum reduction is 10 dB to perceive stationary and natural sounds. The SNR estimation which we recommend applied the principle of correlation/synchrony detection. For the details, refer to the

Appendix in the end.



**Figure 10.** Waveforms of MVDR input, extractor output, and noise remover output in the band 2.5 kHz.

Figure 10 shows performance of a noise remover following the MVDR extractor; these waveforms came from the MVDR beamformer input, the speech extractor output, and the noise remover output. In this experiment, the input voice was selected as that voice in Table 3, except that the incident angle was  $45^\circ$  only; two traffic noises came together, one noise was selected as that noise in Table 3, along the back  $180^\circ$ ; another noise, along the side  $90^\circ$ . Frequency center of the tested bands was 2.5 kHz, and their width was 1 kHz. The two factors  $\alpha$  and  $Z^1$  were optimized under the condition of only one noise coming along  $180^\circ$ . In this figure, the voice waveforms were denoted with blue, the noise waveforms, yellow, and the brown denoted the overlapped voice and noise waveforms. On the top, the voice was drowned by the noise since the voice and noise both were of SPL 60 dB; the mid waveforms resulted from the MVDR extractor output in the band 2.5 kHz. They were amplified by 10 dB to view clearly. The bottom waveforms resulted from the noise remover output, i.e., the waveforms after the silent-interval spectral subtracting; they also were amplified by 10 dB. The silent-interval spectral subtraction rule in Figure 10 is that the noise reduction was divided into three levels: when  $\text{SNR} < -12$  dB, the gain reduced by 9 dB; when  $-12 \text{ dB} \leq \text{SNR} < -9$  dB, reduced by 6 dB; when  $-9 \text{ dB} \leq \text{SNR} < -6$  dB, reduced by 3 dB. Obviously, such a rule almost did not damage the voice waveform. For the different frequency bands, the gain rule may be different for better noise reduction, thus, their reduction level divisions may vary much. By comparing, we can observe that after processing with the silent-interval spectral subtraction, the remaining noise was removed well, especially in the non-voice intervals. However, for some phones, whose SPLs were relatively high, e.g., when  $\text{SNR} > -6$  dB, the noise remained. This is a price to be paid for preserving speech quality. When fitting a hearing aid, an

audiologist can select different gain reduction rules suitable for individual preference in different environments. Through the listening test, the output noise of the noise remover was perceived quite low, and every word could be heard clearly, and be understood fully. The measurement and listening test were conducted with the simulation software, SimuLink R2018b and the audio edition software, Soundbooth.

## 6. Conclusions

In recent years, the development of DM technologies of hearing aids has experienced great challenges, but the hearing aid researchers still made some creative progresses. The research and development mostly conquered some practical weaknesses: the imperfection in low level noise, short suppression to adjacent interference, and failing to perceive multi target voices etc. Accordingly, some novel architectures have been born, such as the front-delay DM, narrow beam DM, and MVDR beamformer. Based on the existing technology references, this paper conducted the extensive researches for implementing architectures and evaluating evidences, finally obtained the new available data and convincing conclusions.

1) The front-delay DM was implemented by exchanging the positions of traditional DM's elements, and replacing the subtracter with an adder. Its speech enhancement was 6 dB, spectrum of its speech output was preserved as in the Omni mic; its performance of noise reduction was better than the Omni mic, but not better than the conventional DM; thus it was suitable for use in a low/soft level noise environment. Simple combination of the front-delay DM with the conventional DM could achieve better noise reduction than that combination of the Omni mic with conventional DM. Siemens' Soft-Level DM featured the front-delay DM.

2) The weighted sum processing of the binaural conventional DMs could implement the narrow beam DM. It enabled the lobe width of the composite beamformer to reduce from  $\pm 90^\circ$  to  $\pm 60^\circ$ , thereby effectively suppressing the interfering noise nearby the target voice. Compared to the conventional DM, it improved SNR by about 4 dB in common practical situations. This technology could cause high-frequency spectrum distortion, which could be decreased by adjusting the weighted coefficients. Phonak's binaural VoiceStream DM featured the narrow beam DM.

3) The MVDR beamformer used Omni mic as the main beamformer, which introduced both multi-azimuth target voices and noises; it used the reverse cardioid DM to estimate the noises. When the main beamformer output subtracted the noise estimator output in the same frequency band of two multiband filter banks, the intruding noises were suppressed/ cancelled. Such processing could detect the open multiple voices effectively, and did not cause speech distortion at all. Additionally, adjusting its delay and block factors could optimally suppress the noises from multiple azimuths. Combined with the silent-interval spectral subtraction, the MVDR beamformer suppressed the surrounding



noises well at competing level with the speech. However, the MVDR beamformer's algorithm is quite complex, and its implementation requires a super high speed processing platform (500 million instructions per second) and very wide dynamic wordlength (24 bits). Velox platform of Oticon can meet such requirements and has been applied to their OSN hearing aids [18].

## Appendix: SNR Estimation Based on Cross-correlation Between Multiple Frequency Bands

### Appendix 1. Correlation Characteristic of Multiband Outputs

In the MVDR noise remover and other noise reduction systems, SNR estimators were required to provide evidential data. Many existing references have discussed SNR estimation algorithms or processors [19], but only a few of them were suitable for the noise reduction of hearing aids. Reference [20] proposed a spectral kurtosis algorithm to estimate SNR of audio signals in their analysis study. Their experimental results showed that the kurtosis of noisy speech could be used to effectively estimate speech and noise energies when the signals were split into narrow bands. The algorithm was based on an assumption of noise model, Gaussian type, which was suitable for mobile telephony. Reference [21] introduced the synchrony detection for hearing aid noise reduction. He illustrated that while the speech energy was distributed across the full bandwidth, the energy patterns in different bands were precisely timed with the periodic action of the vocal folds. His synchrony-detection system demonstrated being sensitive and robust to the speech energy in background noise all the way down to SNR 0 dB and lower.

We conducted further research to provide an available SNR estimation algorithm for the MVDR noise remover. Correlation of a single signal refers to autocorrelation, correlation for short. It usually is a bell-like symmetric function, whose amplitude summit is a positive value with lag=0, lag is a delay time of a signal behind a second signal. Correlation between two different signals refers to cross-correlation. It have an energy summit with lag=0, and its amplitude is positive or negative. Hereinafter, if no lag time is indicated, that means lag=0. In our experiment, some phones /voi/, /e/, /s/, and /pee/ were intercepted from a voice, which was of SPL 60 dB, from the reference [12]; additionally, three sorts of noises, traffic, white, and babble, were acquired from the references [13, 14, 22], and calibrated SPL 60 dB. Eight equal bandwidth FIR filters were employed; the center frequencies of the filters were 0.5, 1.5, ..., 7.5 kHz, and their bandwidths all were 1 kHz [4]. In fact, a voice signal is composed of some fundamental waves and their harmonic components, so the voice signal samplings are of a certain correlation. Furthermore, when a voice signal went through the test filter bank, the waveforms of the eight bands' outputs were cross-correlated precisely. We intercepted the short voice

30 ms to clearly view the waveforms, as showed in Figure A1 with a phone /voi/, because the compact waveforms of voice signals looked disordered. It can be observed that the waveforms were synchronous or turnover synchronous, especially the waveforms from two adjacent bands. Of the eight bands, the waveforms from three bands 1.5, 6.5, and 7.5 kHz were turnover synchronous with those from the other bands; the cross-correlation values between the waveform and turnover waveform were negative.

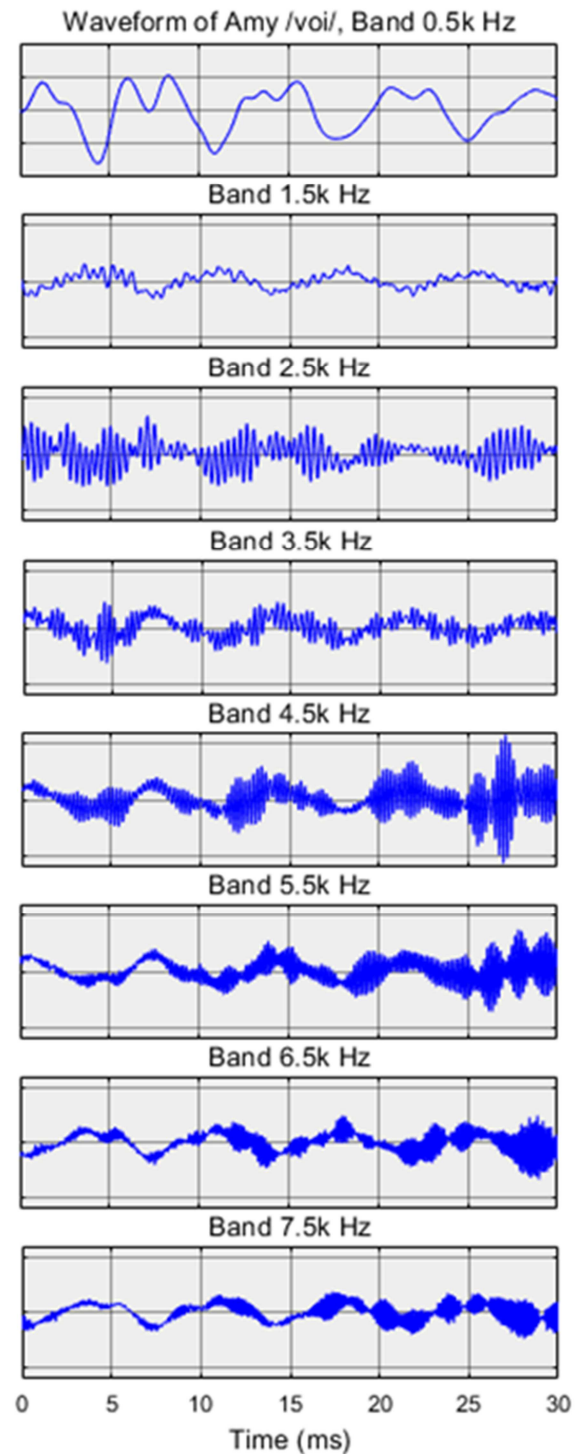
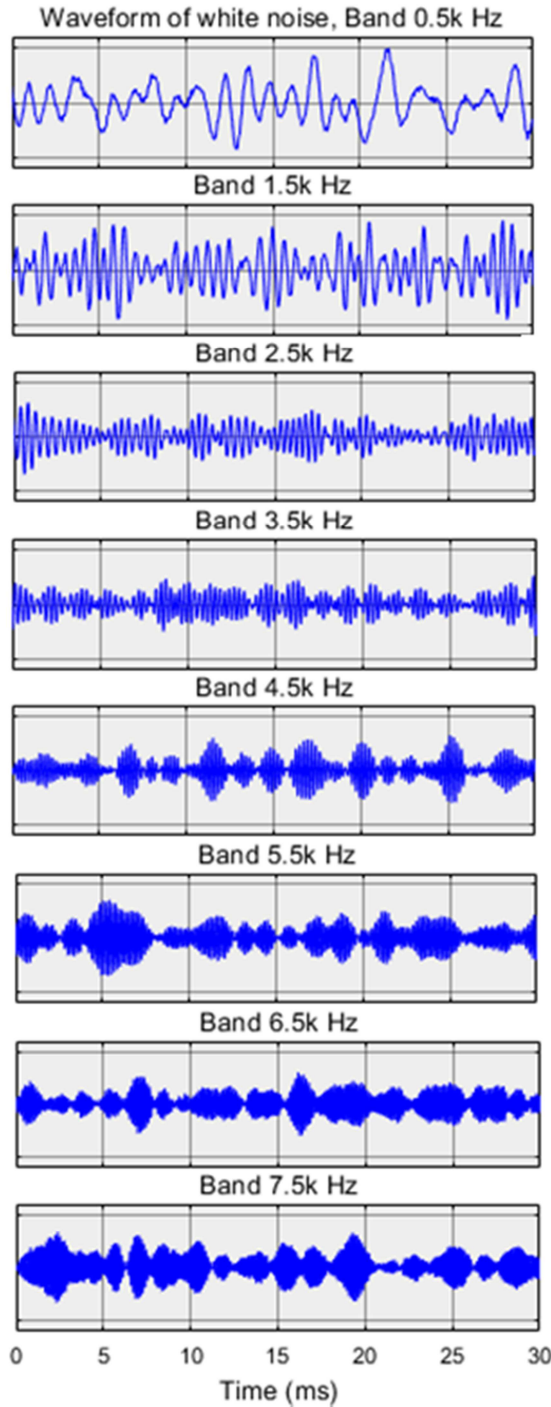


Figure A1. Waveforms of multiband outputs with a phone /voi/.





**Figure A2.** Waveforms of multiband outputs with a white noise.

Usually, a noise is caused by some physical phenomena, e.g., weather, traffic, equipment and party etc., instead of the vocal fold vibration. Thus, the noise signal moves at any time, and its amplitude always is random. Generally, samplings of the noise waveform are independent; its autocorrelation looks like a sharp pole, the correlation values always drop rapidly with the lag  $\neq 0$ . The cross-correlation between two adjacent band outputs with lag=0 can be positive or negative, and its energy drops rapidly with the lag increase. For example, when the test white noise went through the above multiband filters, their output waveforms were showed in Figure A2, and the

view period was 30 ms too. The noise was from a 4 s time-series of SPL 60 dB. Obviously, the waveforms did not show any synchrony behaviors. During calculating the cross-correlation between two band outputs, we ascertained that the correlation values were affected by the output power. To avoid the impact of the output power on the correction values, the further experiments were done. Two available solutions to power impact were found: to normalize the outputs' RMS and to apply zero-cross function to the outputs. The latter was simpler and more effective than the former. Additionally, the cross-correlation of either the noises or speeches would decrease with lag increase, but that of the noises decreased faster than that of the speeches. So, setting the lag a few sampling periods made a larger difference between the two correlations of the voices and the noises than setting the lag zero. When the number of the multiband filters was increased, e.g., 16, the cross-correlations between the two adjacent bands' outputs (with the speech or noise) increased insignificantly.

## Appendix 2. Measurement Results

Table A1 lists four columns of the cross-correlation values between the in-pair bands' outputs vs. input SNRs of the phone /voi/ plus the traffic noise. We set the lag 2 sampling periods in this and late tests. The input with the test phone plus the noise lasted 0.259 s to cover the full /voi/ duration. When  $\text{SNR} = +\infty$  dB, the input was the phone only; when  $\text{SNR} = -\infty$  dB, the input was the noise only. We can observe that (1) when the input SNRs varied from high to low,  $+\infty, 0, -3, -6, -9, -12$  to  $-\infty$  dB, the obtained correlation values also went from high to low, e.g., for the in-pair bands 2.5 and 3.5 kHz, the corresponding correlation values were 0.566, 0.319, 0.250, 0.184, 0.134, 0.099 to 0.058. (2) When the input was /voi/ only, the correlation values of the four in-pair bands were the highest. (3) When the input was the traffic noise only, the correlation values of the four in-pair bands were the lowest.

**Table A1.** SNRs vs. correlation values of the in-pair bands' outputs with phone /voi/ plus traffic noise.

SNRs (dB)	Correlation values of phone /voi/ plus traffic noise from in-pair bands (kHz)			
	0.5&1.5	2.5&3.5	4.5&5.5	6.5&7.5
$+\infty$	-0.292	0.566	0.784	0.731
0	-0.109	0.319	0.600	0.630
-3	-0.083	0.250	0.503	0.565
-6	-0.056	0.184	0.406	0.492
-9	-0.024	0.134	0.315	0.422
-12	-0.019	0.099	0.244	0.362
$-\infty$	-0.010	0.058	0.154	0.242

Table A2 lists four columns of the cross-correlation values between the in-pair bands' outputs vs. input SNRs of the phone /voi/ plus the babble noise. The test conditions were the same as those in Table A1. We can observe that (1) when the SNRs varied from high to low,  $+\infty, 0, -3, -6, -9, -12$  to  $-\infty$  dB, the obtained correlation values also went from high to low, e.g., for the in-pair bands 4.5 and 5.5 kHz, the corresponding correlation values were 0.784, 0.537, 0.439, 0.332, 0.264,

0.219 to 0.164. (2) When the input was /voi/ only, the correlation values were the same as those in Table A1. (3) When the input was the babble noise only, the correlation values were the lowest, and different from those in the traffic noise. Overall, these correlation values were close to those in Table A1.

**Table A2.** SNRs vs. correlation values of the in-pair bands' outputs with phone /voi/ plus babble noise.

SNRs (dB)	Correlation values of phone /voi/ plus babble noise from in-pair bands (kHz)			
	0.5&1.5	2.5&3.5	4.5&5.5	6.5&7.5
+∞	-0.292	0.566	0.784	0.731
0	-0.129	0.371	0.537	0.619
-3	-0.097	0.292	0.439	0.544
-6	-0.078	0.232	0.332	0.472
-9	-0.046	0.184	0.264	0.417
-12	-0.019	0.151	0.219	0.369
-∞	-0.007	0.093	0.164	0.310

Table A3 lists four columns of the cross-correlation values between the in-pair bands' outputs vs. input SNRs of the phone /voi/ plus the white noise. The test conditions were the same as those in Table A1. We can observe that (1) when SNRs varied from high to low, +∞, 0, -3, -6, -9, -12 to -∞ dB, the obtained correlation value also went from high to low, e.g., for the in-pair bands 6.5 and 7.5 kHz, the corresponding correlation values were 0.731, 0.109, 0.065, 0.037, 0.020, 0.012 to 0.003. (2) When the input was /voi/ only, the correlation values were the same as those in Table A1. (3) When the input was the white noise only, the correlation values were the lowest. Overall, these correlation values in Table A3 were lower than those in Tables A1 and A2, except for two values in the in-pair bands 0.5 and 1.5 kHz. This indicated that independence of the white noise was stronger than that of the traffic and babble noises.

**Table A3.** SNRs vs. correlation values of the in-pair bands' outputs with phone /voi/ plus white noise.

SNRs (dB)	Correlation values of phone /voi/ plus white noise from in-pair bands (kHz)			
	0.5&1.5	2.5&3.5	4.5&5.5	6.5&7.5
+∞	-0.292	0.566	0.784	0.731
0	-0.110	0.091	0.124	0.109
-3	-0.081	0.062	0.068	0.065
-6	-0.050	0.037	0.043	0.037
-9	-0.037	0.023	0.028	0.020
-12	-0.025	0.018	0.021	0.012
-∞	-0.018	0.010	0.016	0.003

According to the correlation values in Tables A1, A2, and A3, this SNR estimation algorithm can be provided for hearing aid noise reduction as a detector. For the MVDR noise remover application, the correlation values in these Tables can be used in the case of eight bands. In addition, we suggest that (1) for a practical SNR estimator, it is necessary to further test this algorithm under some music conditions. The music signals have different correlation values from the voices. (2) In particular calculations, the table of SNRs vs. correlation values is related to not only the noise sorts but also correlation

window length etc., so further experiments should be essential.

## References

- [1] Keidser, G., Rohrseitz, K., Dillon, H., Hamacher, V., Carter, L., Rass, U., Convery, E., The effect of multi-channel wide dynamic range compression, noise reduction, and the directional microphone on horizontal performance in hearing aid wearers. *Int. J. Audiol.*, 2006, 45 (10), 563-579.
- [2] Flynn, M. C., Maximizing the voice-to-noise ratio (VNR) via voice priority processing. *The Hearing Review*, 2004, 11 (4), 54-59.
- [3] Moeller, K., Jespersen, C., The effect of bandsplit directionality on speech recognition and noise perception. *The Hearing Review*, 2013, 20 (5), 17-24.
- [4] Zhang, X., Spectrum distortion of a directional microphone and its removal for hearing. *IJISMS*, 2019, 3 (5), 14-22.
- [5] Phonak Insight Paper, Binaural directionality. Stafa, Switzerland, 2010, pp. 1-4.
- [6] Chalupper, J., Wu, Y. H., Improving effectiveness of directional microphones with Soft Level Directivity. *The Hearing Journal*, 2011, 64 (7), 44-46.
- [7] Latzel, M., Binaural VoiceStream Technology® - Intelligent binaural algorithms to improve speech understanding. Phonak Insight, 2012, Stafa, Switzerland, V1: 1-3. [https://www.phonakpro.com/content/dam/phonakpro/gc\\_hq/en/resources/evidence/white\\_paper/documents/Insight\\_Binaural\\_VoiceStream\\_Technology\\_028-0773.pdf](https://www.phonakpro.com/content/dam/phonakpro/gc_hq/en/resources/evidence/white_paper/documents/Insight_Binaural_VoiceStream_Technology_028-0773.pdf).
- [8] Kamkar-Parsi, H., Pipl-Ing, E. F., Aubreville, M. New binaural strategies for enhanced hearing. *The Hearing Review*, 2014, 21 (10), 42-45.
- [9] Vorobyov, S. A., Principles of minimum variance robust adaptive beamforming design. *Signal Processing*, 2013, 93 (2013), 3264-3277.
- [10] Le Goff, N., Jensen, J., Pedersen, M. S., Callaway, S. L., An introduction to OpenSound Navigator™. White Paper, 2016, Oticon, A/S.
- [11] Application Note, Directional microphones. 2019, Copyright by Sonion. <https://www.sonion.com/hearing/microphones/directional-microphones/>.
- [12] Voice Samples-LAMP, Free voice for words with life. 2020, Copyright by PRC-Saltllo. <https://aacapps.com/lamp/voices>.
- [13] Traffic Wav & Mp 3, Traffic sound effects. 2019, Copyright by ZAPSPLAT. <http://www.pachd.com/traffic-ambience.html>.
- [14] MathWorks, Band-limited white noise/Sources/ Simulink. Matlab, R2018b, 2018.
- [15] Zhang, X., Benefits and limitations of common directional microphones in real-world sounds. *Clinical Medicine Research*, 2018, 7 (5), 103-118. Doi: 10.11648/j.cmr.20180705.12.
- [16] Suchum, D. J., Noise reduction via signal processing: (1) Strategies used in other industries. *The Hearing Journal*, 2003, 56 (5), 27-32.

- [17] Kahrs, M., Brandenburg, K., Applications of Digital Signal Processing to Audio and Acoustics. Kluwer Academic Publishers, ©2002, New York, USA.
- [18] Weile, J. N., Bach R., The Velox™ platform. Tech Paper, 2016, Oticon A/S.  
<https://www.oticon.global/professionals/training-and-fitting/training/e-learning/velox-platform>.
- [19] Cracium, A., Gabrea, M., Correlation coefficient-based voice activity detector algorithm. Canadian Conference on ECE, 2004, 3: 1789-1792. Doi: 10.1109/CCECE.2004.1349763.
- [20] Nemer, E., Goubran, R., SNR estimation of speech signals using subbands and four-order statistics. IEEE Signal Processing Letters, 1999, 6 (7), 171-174.
- [21] Suchum, D. J., Noise-reduction circuitry in hearing aids: (2) Goals and current strategies. The Hearing Journal, 2003, 56 (6), 32-41.
- [22] Fontan, L., Tardieu, J., Gaillard, P., Woisard, V., Ruiz, R., Relationship between speech intelligibility and speech

comprehension in babble noise. J. Speech Lang Hear Res., 2015, 58, 977-986. Doi: 10.1044/2015\_JSLHR-H-13-0335.

## Biography



Xubao Zhang received his doctorate in electronics from Xi'an Electronic Science and Technology University in China and was a postdoctoral fellow at McMaster University in Canada. He has been interested in hearing aid technology strategies and performance evaluation. He worked as EA and EMC engineer with Sonova Unitron in Canada, and also worked with Oticon Canada and Beltone Canada. And he worked as associate professor at EE department of this University in China for radar signal processing research. He is the author of one book and more than 40 articles.