Designs, Experiments, and Applications of Multichannel Structures for Hearing Aids

Xubao Zhang

Abstract — The structures of common multichannel processing for hearing aids include equal bandwidth (BW) finite impulse response (FIR) filter bank, nonuniform BW FIR filter bank, and fast Fourier transform (FFT) plus inverse FFT (IFFT). This paper analyzes their operation principles, indicates the design methods by means of MATLAB R2018b resources, and describes the main characteristics: synthetical ripple, bank filters' group delays, and individual filter sidelobe attenuation. Three schemes are proposed: equal BW sixteen-filter bank, logarithmic BW eight-filter bank, and 128-point FFT plus IFFT with overlap-add operation. To build the experimental modules, we introduce the settings of spectrum scopes, the acquirement of realistic speech and noises, and the gain enhancing/reducing needs of hearing aid features; the characteristics of synthetical outputs confirm precise control ability of the multichannel modules and differences between the three schemes. Subsequently, this paper illustrates two applications of the multichannel structures in hearing aids, the equal BW sixteen-filter bank with spectral subtraction (SS) for an artificial intelligence (AI) noise reduction (NR) and 128-point FFT plus IFFT spectral distortion removal for a directional microphone (DM). In Amy's speech mixed with ringing, milk steamer, and strong wind noises separately, the SS processor improves signal-noise-ratio (SNR) by 6.5 to 15.9 dB. By measuring waveforms and spectra at the DM input and output, the DM system seamlessly removes the spectral distortion.

Index Terms — equal bandwidth filter, FFT, IFFT, multichannel structure, octave bandwidth filter, overlap-add.

I. INTRODUCTION

Multichannel processors have been applied to a variety of signal processing areas, such as radar return processing, speech signal processing, and hearing-aid signal processing, for several decades. A quality multichannel structure can achieve refined gain response control in a desired frequency region. Reference [1] researched broadband radar return processing, using the minimum variance distortionless response (MVDR) technique, which joints spatial-domain and frequency-domain processing so that the broadband return is split into narrow processing signals by FFT, then the signals are processed in multichannel beamformers. After interferences are suppressed, the beamformers' outputs are synthesized by IFFT to achieve robust adaptive MVDR processing. In speech recognition research, both noise reduction (NR) and speech recognition stages with multiple channels provide more details than with a single channel. References [2], [3] introduced the philosophies of speech

recognition; they selected discrete Fourier transform (DFT) as the multichannel scheme rather than a bandpass-filter bank due to more dimensionality and smaller calculation load. These multichannel processors gained benefits of the additional information existing in multiple observations.

Nowadays, conventional dynamic range compression (DRC) and frequency warping DRC both incorporate the multichannel processing. Reference [4] used an all-pass-filter bank, which gives Bark bands with the low-frequency bands approximately spaced at multiples of 135 Hz and with the high-frequency bands, 1.8k Hz, and added a side branch of FFT plus IFFT operation to analyze the intensity of input noisy speech and to provide the warped filters with assigned gains to run their multichannel compression. Reference [5] proposed a SNR-aware DRC, which is composed of short-term Fourier transform (STFT) plus inverse STFT (ISTFT) operators and a dynamic gain estimator and relies not only on the input noisy speech intensity but also on the input SNR. They designed an FIR bandpass-filter bank of seven octave bands with center frequencies 125, 250, 500, 1k, 2k, 4k, 8k Hz; both noisy speech power and speech power spectral density from the STFT output are applied to the bank to estimate short-time SNRs. The SNR decision is utilized to adaptively adjust release time of the compressor. Reference [6] described the features of Oticon MoreSound Intelligence™, whose Neural Clarity processing adopts a 24-filter bank (an FFT bin is also an FIR filter.) to finely analyze the incoming signals and to acquire the noise intensity and SNR in each channel; two different processing ways are to handle the speech-dominated and noise-dominated incoming signals separately; in the 24-channel SS way, a channel's gain reduction depends on its noise intensity and SNR; then the two-way outputs are prioritized to synthetically output quality speech. This is referred to as Neural Clarity Process due to similarity to the human neural activities via environment training. Reference [7] described the hardware (HW) platform named Polaris for the MoreSound products. Compared to its last generation, Polaris can embed more intelligent activity instructions and store more priori and learning knowledge. Its triple multichannel structures should be to adopt FFT plus IFFT schemes, and their filter BWs are nonuniform for fitting speech-spectrum details, but are not octave, not Bark scale either.

Based on the above achievements, this paper analyzes multichannel structures for hearing aids and proposes three typical schemes, including equal BW sixteen-filter bank,

---

Submitted on July 13, 2021.

Published on August 03, 2021.

Xubao Zhang, Sonova Unitron, Kitchener, Canada; Xi'an Electronic Science and Technology University, Xi'an, China.
(corresponding e-mail: xbzwdl@yahoo.com)

logarithmic (Log) BW eight-filter bank, and 128-point FFT plus IFFT with overlap-add operation; we also make experiments to evaluate the effectiveness of the multichannel schemes and to compare their characteristics. Subsequently, we build two application modules, a multichannel SS for AI noise reduction (NR) and a multichannel gain balancer for directional microphone (DM) spectral distortion removal and disclose the test results.

II. SCHEMES AND DESIGNS OF MULTICHANNEL STRUCTURES

In the times of analog signal processing, a hearing aid circuit always worked with a single channel processor. A few decades ago, multichannel processors came up along with the birth of digital hearing aids [8]. Because of digital processing flexibility and refinement, a multichannel processor can achieve high-level features, such as digital NR, wide DRC, multichannel DM, etc. Multichannel is also referred to as band-split. In a multichannel processor, when an incoming signal is split into N_c channels, the anatomy of signals for learning is referred to as analysis. After feature processing and decision making, the multichannel sub-signals are summed, and the summation is referred to as synthesis. A multichannel structure is required to cover the full frequency range, from 100 to 8k Hz or wider, to meet the relevant requirements of ANSI and IEC hearing aid standards. Two or more features can be carried out in one multichannel scheme. Common multichannel structures have three schemes.

A. Equal Bandwidth Filter Bank

The scheme of equal BW FIR filter bank incorporates multiple, parallel, equal BW bandpass FIR filters to ensure audio coverage. The larger the number of bandpass filters N_c is, the narrower the filter BWs are, and the more finely the processing is conducted, but the heavier the calculation load and power consumption are. In Fig. 1, blocks of a multichannel scheme of equal BW filter bank, $N_c=16$, are shown, denoted by EB16; the dashed box contains various possible features. The filter length N_p and sampling rate F_s affect the frequency response and group delay. So, determining N_c , N_p , and F_s is the first step to design the filter bank. The frequency response of the filter bank EB16 covers 100 to 8k Hz; BW of each filter is 500 Hz except that the 1st BW is 400 Hz; their center frequencies are 300, 750, 1.25k, 1.75k, ..., 7.25k, 7.75k Hz. The second step is to design all the bandpass filters, a heavy work. The design requirements are that ① the ripples of synthetical output are as low as possible; ② the group delays of all filters are as equal and short as possible; ③ the passband gain of each filter is 1 as possible. It is efficient to utilize Simulink of MATLAB to design the filters. In its block Bandpass Filter, our basic settings were as follows: Sampling rate F_s , 44.1k Hz (conventional in audio signals); Structure, Direct-Form FIR; and Algorithm, Equal-ripple. After all the individual filters were designed well, we had to examine the frequency response of the synthetical output because all the individual

optimized filters cannot ensure the filter bank to be optimal. Usually, it is necessary to repeatedly adjust all the individual filters' designs until the synthetical output is optimized. Fig. 2 shows two frequency responses: one results from the all-pass filter bank (all the filters are working with gain 1), black curve and the other from a working filter, center frequency 1.75k Hz, blue curve. By measuring, we obtained that the intensity of the all-pass bank output is 9.2 dB while the signal source output is 9.09 dB, and the ripples are within ± 0.8 dB; the 1st sidelobe attenuation of the filter 1.75k Hz is about 20 dB* and the transition band width is about 200 Hz, narrow enough. These characteristics are quite good, but the price is the filter length $N_p=513$ and group delay 5.8 ms; the filter bank can be redesigned if we require smaller group delay. However, when the group delay is designed to be smaller, the filter length has to be shorter, then the ripples get higher. For the measuring method, see subsection III.A. In HW implementation, the equal BW FIR filter bank needs $N_c N_p/2$ multiplications and $N_c N_p/2$ additions.

Fig. 1. Blocks of an equal BW filter-bank scheme EB16.

Fig. 2. Frequency responses of multichannel scheme EB16 with all-pass bank and an individual filter.

B. Logarithmic Bandwidth Filter Bank

Audiology research found that the frequency resolution of human hearing decreases with frequency increase [4]; in addition, the speech energy is distributed across all the voice frequencies but most of the energy is concentrated in fundamental and harmonic components, i.e., in the low and mid frequency regions. It is reasonable to more finely analyze the speech spectral components at low and mid frequencies. As a result, some manufacturers preferred nonuniform BW filter banks for their multichannel schemes. This can have two common types: Log BW and Bark band. The former is selected as our scheme here. The Log BW of a band is

*The unit of the magnitude axis in the spectral figures is dBm/Hz, we replace it with dB in this text for simplicity.

referred to as the octave BW, which is proportional to its center frequency. Fig. 3 shows blocks of a Log BW FIR filter-bank scheme, $N_c=8$, denoted by LB8. After deciding the number of bank filters N_c , the filter length N_p , and the sampling rate F_s , we can design the filter bank. Its design method is the same as that used in the EB16 except that the BWs are following the 2/3 octave band scale. We can easily search out on line that these BWs are [200, 315], [315, 500], [500, 800], [800, 1.25k], [1.25k, 2k], [2k, 3.15k], [3.15k, 5k], and [5k, 8k] Hz; their center frequencies are 250, 400, 630, 1k, 1.6k, 2.5k, 4k, and 6.3k Hz, respectively. The frequency response of the filter bank covers the frequency range of 200 to 8k Hz. For the lower frequency coverage, we can add one filter, which has center frequency 160 Hz and BW [125, 200] Hz. Since these filter BWs are much different, from 115 to 3k Hz, the resulting filter lengths are also much different, from 800 to 20. Fig. 4 shows two resulting frequency responses, one results from the all-pass filter bank and the other from a working filter, center frequency 2.5k Hz. This scheme uses half number of bank filters of the scheme EB16 but also covers the desired frequency range. By measuring, the intensity of the all-pass filter bank output is around 9.3 dB and the ripples are within ± 1.5 dB; 1st sidelobe attenuation of the filter 2.5k Hz is 21 dB and the transition band widths are 273 (left) and 409 (right) Hz. The individual filters hold quite different group delays, from 0.374 to 9.25 ms. The wider the filter BW is, the shorter the filter length and delay are. Most characteristics of the scheme LB8 are not better than those of the scheme EB16. The large delay differences can suffer from severe delay distortion, causing abnormal waveform of the synthetical output. However, our listening test confirmed that the synthetical speech sound did not change obviously as the synthetical waveform did. The Log BW filter bank has the same formula of calculation load and the load of the scheme LB8 is much lighter than that of the scheme EB16 due to much smaller N_c .

Fig. 3. Blocks of a Log BW FIR filter-bank scheme LB8.

Fig. 4. Frequency responses of multichannel scheme LB8 with all-pass bank and an individual filter.

C. FFT plus IFFT with Overlap-Add Operation

FFT and IFFT partition DFT and inverse DFT (IDFT), respectively, to radix-2 butterfly operation units, which carry out the same function with high efficiency [9]. For convenience, we still use the DFT and IDFT principles to describe an FFT plus IFFT multichannel scheme. When an input signal $x_s(n)$, $n \in \{0, 1, \dots, N_t-1\}$, N_t the signal length, enters a DFT operator, which is directly followed by an IDFT operator, the IDFT output $y_s(n) = x_s(n)$ exactly. This is so-called DFT reconstruction. The DFT of $X_s(k)$ can be calculated by:

$$X_s(k) = \sum_{n=0}^{N_t-1} x_s(n) e^{-j2\pi nk/N_t} \quad k \in \{0, 1, \dots, N_t-1\}, \quad (1)$$

where N_{ft} is the DFT length. In practical application, we always multiply $\{X_s(k)\}$ by a gain response $\{H(k)\}$, which is the DFT of a filter impulse desired by a hearing aid feature, $h(l)$, $l \in \{0, 1, \dots, N_h-1\}$, N_h the impulse length; the resulting output $y_s(n)$ is the IDFT of $\{X(k)H(k)\}$. In order to fit the DFT operation rule, a buffer is needed to buffer a sample-based input to frame-based input. Generally, the IDFT output is a circle convolution of $x_s(n)$ and $h(l)$, rather than the conventional linear convolution,

$$y_s(n) = \sum_{l=0}^{N_h-1} x_s(n-l)h(l). \quad (2)$$

If the DFT length N_{ft} is equal to or larger than N_t+N_h-1 , the DFT plus IDFT output is equal to the linear convolution of (2). In practice, N_{ft} is always limited and N_t is very long, even unlimited, so this assumption is hardly true. As a result, the temporal aliasing always occurs when cascading the adjacent frames' data of the IDFT output. Alternatively, the input signal is buffered to sequential framed-samples of length N_b and N_h-1 zeros are padded at the back end of the samples so that the framed input length is extended to the DFT length $N_{ft}=N_b+N_h-1$. Then, the fore N_b data of IDFT adjacent frame outputs are cascaded together to produce the sample-based $y_s(n)$ without aliasing. The solution of the padding zeros and cascading framed outputs is so-called the overlap-add method [9]. In addition, the STFT plus ISTFT multichannel structure also behaves with similar function to the FFT plus IFFT with overlap-add operation [10], but it utilizes a windowing operation rather than overlap-add operation to eliminate the temporal aliasing.

Based on the principles of above FFT plus IFFT multichannel structure, we designed a scheme by means of Simulink. Fig. 5 shows its blocks: buffer, Zeros pad, FFT, Frequency gain response desired by hearing aid features, IFFT, and Selection and Add; the overlap-add operation is carried out by the Zeros pad and Selection and Add. The FFT length is 128, the sampling rate is 44.1k Hz, and the bin BW is 344.53 Hz. The bin factors of FFT and IFFT operators are inherent in their blocks, so we do not need to design; in HW implementation, they can be found from any DSP (digital signal processing) book. Because of real signals of hearing aid circuits, the FFT outputs of bins 1 to 64 only are multiplied by a real gain response; the IFFT setting is 64-point conjugate symmetric input. Furthermore, a bank of fore 24 bins is enough to cover a frequency range of 0 to 8.1k Hz, so the FFT output channels of the bins 25 to 64 do nothing,

independent of the features. Fig. 6 shows two measured frequency responses, one results from the scheme with the gain response $H(k)=1$ at bins 1 to 64, black curve, and the other from the gain response 1 at two working bins 5.2k and 5.5k Hz (channels 16 and 17) and gains 0 at the rest of all bins, blue curve. By measuring, we obtained that the black response intensity is 9.09 dB and the ripples are within ± 0.09 dB; the blue response has a center frequency 5.34k Hz and BW 689Hz; the attenuation of 1st sidelobe is 13.3 dB, not deep. The all-pass response of FFT plus IFFT scheme are very smooth and the ripples are extremely low; the response of two bins looks quite regular and matches the theoretical response. For measurement methods, see subsection III.A. In comparison to the schemes EB16 and LB8, the IFFT synthesis does not cause delay differences between the channels but it has a shortcoming of higher sidelobes. The FFT plus IFFT scheme needs $N_{ft} \log_2 N_{ft} - 1.5 N_{ft} + 4$ real multiplications and $1.5 N_{ft} \log_2 N_{ft} - 1.5 N_{ft} + 2$ real additions. Note that generally FFT and IFFT operate in a complex mode but the signal processing of hearing aids is in a real mode. In HW implementation, one should adopt fast discrete cosine transform (DCT) operation for higher efficiency [11], so the calculation load that we give here is from the fast DCT.

Fig. 5. Blocks of an FFT plus IFFT scheme with overlap-add operation.

Fig. 6. Frequency responses of FFT plus IFFT scheme with gains 1 at all bins and gains 1 at two working bins.

By comparing the above schemes, we conclude that ① in the two filter-bank schemes, the filter BW depends on the desired frequency range and number of the bank filters; but in the FFT plus IFFT scheme, the filter (bin) BW depends on number of the FFT bins and the sampling rate; usually, it is not necessary to use all the FFT bins to cover the desired frequency range. ② The FFT plus IFFT operation is equivalent to a linear convolution of the input signal with a filter desired by the features; in fact, the filter is absent but its DFT is present, just the gain response. ③ The synthesis operation of filter-bank schemes simply runs summing of all

channels' outputs, probably causing delay distortion, but the synthesis of FFT plus IFFT scheme adopts the IFFT operation without such a distortion problem. ④ All the characteristics of the filter bank schemes can be assigned but those of the FFT plus IFFT scheme cannot except the bin BW. ⑤ The efficiency of FFT plus IFFT is higher than that of the filter banks. For example, number of multiplications are 4104 in the scheme EB16 and 708 in the scheme FFT plus IFFT.

III. EXPERIMENTS OF THE MULTICHANNEL SCHEMES

After a multichannel scheme is designed well, we need to test its basic performances. The hearing-aid multichannel processing always incorporates some features: NR, DRC, feedback (FB) cancellation, and DM gain balancer; either of them is carried out by enhancing and/or reducing gains of the channels.

A. Experimental Preparations

We can conveniently get sine wave signals, waveform scopes, and spectrum scopes in Library of Simulink. In order to get reliable measures, we first build a highly accurate signal source that meets our experimental requirements. Particularly, its output spectrum has extremely low ripples and covers the frequency range required by ANSI and IEC hearing aid standards. The built source consists of 129 sine waves, intervals of the adjacent wave frequencies are 62.5 Hz, and the frequency coverage is from 2 to 8k Hz. Fig. 7 shows the spectrum of signal source output; the intensity is 9.09 dB and the ripples are within ± 0.09 dB. This is the highest-quality source that we have ever built.

When we use the block Spectrum Scope, different parameter settings may result in different measuring results, i.e., inconsistency issue. We set the main parameters of Spectrum Scope as ① the view time, 0.5 s due to the lowest frequency 2 Hz of the source; ② the length of analytical FFT inside the scope, 512; ③ the window function, rectangle in order to view unsmoothed spectrum; ④ FFT estimating average times max, $44,100\text{Hz} \times 0.5\text{s} / 512 \approx 43$. When viewing a speech spectrum on the scope, it is necessary to set the maximum average times because every frame-based estimate is a different short-term spectrum. For example, Amy's voice "Hi, one of the available high-quality texts to speech voices" is of duration 3.83 s and from a speech wave file [12]. Fig. 8 shows three measured spectra on Spectrum Scope with three average times: 330 (Amy's 11 words), 220 (8 words), and 110 (4 words); obviously, the spectra differ greatly, especially in the mid and high-frequency regions.

Fig. 7. Spectrum of a lab-built signal source output.

Fig. 8. Speech spectra measured with three different average time.

B. Experiments with Three Multichannel Schemes

Simulink Library and DSP System Toolbox of MATLAB provide many functional blocks, which are enough to carry out the expected experiments. Their signal source, test scopes, multipliers, adder, FIR filters, FFT, and IFFT, etc. can be connected together to form desired experimental modules so as to conduct the behavior evaluation, data acquirement, and important results' storage.

1. Based on the scheme EB16 in subsection II.A, we built an experimental module as shown in Fig. 9. All the Bandpass blocks' filters were designed in subsection II.A. The blocks V-Sum and Adder are vector and sample summers, respectively. The block Sine vectors plus V-Sum is the source built in subsection III.A; the blocks Input waveform and Output waveform are waveform scopes; the blocks Input spectrum and Output spectrum are spectrum scopes. There are optional enhancing and reducing gains denoted by multipliers. Assume that some feature requires that the gain of multiplier M3 is 2 at channel 2.25k Hz, the gain of M6 is 0.125 at channel 4.75k Hz, and the gain of M7 is 0.25 at channel 5.25k Hz. Fig. 10 shows a frequency response of the module EB16 with the above multipliers. We observe that a bell-like curve appears around 2.25k Hz and a notch curve around 4.75k to 5.25k Hz; their transition bands are quite narrow. This response confirms our expectation: 6 dB gain enhancing with 500 Hz BW and -18 dB to -12 dB gain reducing with 1k Hz BW. Note that the left notch is not 18 dB deep due to the effects of left, adjacent filter sidelobes. Such characteristics are essential for various features.

Fig. 9. An experimental module of multichannel scheme EB16.

Fig. 10. A frequency response of experimental module EB16.

2. Based on the scheme LB8 in subsection II.B, we built an experimental module as shown in Fig. 11. All the blocks are the same as those in Fig. 9 except for the Bandpass filters and multipliers. The eight bandpass filters hold 2/3 octave BWs and their frequency coverage is from 200 to 8k Hz. In this experiment, assume that some feature assigned the enhancing gain 2 to multiplier M5 at channel 1.6k Hz and the reducing gain 0.25 to M7 at channel 4k Hz. To confirm the effectiveness of gain control, the frequency response of this module was measured as shown in Fig. 12. We observe that a bell-like curve appears around frequency 1.6k Hz, 6 dB high, and a notch curve around frequency 4k Hz, 12 dB deep. This response reaches what we expected of the feature's need. In comparison to the module EB16, the filter BWs of the module LB8 are quite narrow in the low frequency region, beneficial to finely analyze speech spectrum; the filter BWs in the high-frequency region are quite wide, enabling number of the bank filters to lessen; the transition bands are wider than those of the module EB16.

Fig. 11. An experimental module of multichannel scheme LB8.

Fig. 12. A frequency response of experimental module LB8.

3. Referring to Fig. 5, we know that to build an experimental module of the FFT plus IFFT scheme, we need to incorporate many functional blocks and to set many parameters inside them. Alternatively, there is a preferable block in DSP System Toolbox named Frequency-Domain FIR Filter, which contains all the blocks in Fig. 5. Based on the design in subsection II.C, we built a compact FFT plus IFFT experimental module, using the core block Frequency-Domain FIR Filter and a gain table Frequency response as shown in Fig. 13. Many parameters of all the subblocks are integrated; our main settings inside are as follows: Frequency-domain filter method, Overlap-add operation; Numerator domain, Frequency; Specify frequency response from input, checked; Time-domain numerator length, 64; Filter is real, checked; and Output filter latency, checked. All the scopes are the same as those in Fig. 9. What we need to emphasize is that ① Frequency-Domain FIR Filter has an inherent complex mode, (Blocks FFT and IFFT have a real mode option.) so the table Frequency response is symmetrically filled out; ② the gain table Frequency response is of FFT length 128 and the table data are provided externally for accuracy and simplicity; ③ the length of potential FIR filter is necessarily smaller than or equal to half of the FFT length, otherwise the IFFT output distorts. When the Frequency response table is a unit vector of length 128, the output response is a transcription of the signal source spectrum as the black curve shown in Fig. 6. This confirms the reconstruction characteristic of FFT plus IFFT. Assume that some feature requires that the table Frequency response has an enhancing gain 4 at bins 5 and 6 and reducing gain 0.125 at bins 17 and 18. The measured frequency response of this module is shown in Fig. 14. We observe that a bell-like curve, 12 dB high, and a notch curve, 18 dB deep, appear at the corresponding frequency regions 1.72k to 2.07k Hz and 5.86k to 6.2k Hz, respectively; the transition bands look relatively wide. Note that the -18 dB reducing gain need results in the notch of around -13 dB only, this is due to the effects of high, adjacent bin sidelobes, referring to Fig.6. When the FFT and IFFT lengths increase, this problem can be relieved.

Fig. 13. An experimental module of 128-point FFT plus IFFT scheme.

Fig. 14. A frequency response of 128-point FFT plus IFFT scheme with a frequency response table.

IV. A MULTICHANNEL SCHEME FOR NOISE REDUCTION

When an individual with hearing loss listens to a voice, he perceives not only less audible but also blurry understood speech, especially to a voice with noise. Furthermore, voice and noise signals sometimes mix together in the whole speech spectrum. These facts make a great challenge to hearing aid designers [13], [14]. Nevertheless, differences between voice and noise signals still exist and can be utilized. A voice results from vibration of an individual vocal fold and is composed of its fundamental and harmonic waves fluctuating rapidly in amplitude and spectrum. Generally, a noise is caused by some physical event, such as traffic, cocktail party, weather, etc.; its amplitude fluctuation is not that rapid, and its spectrum is relatively steady and narrow. Referring to the hearing-aid NR developments [6], [13], and [14], we built a multichannel spectral subtraction (SS) module for artificial intelligence (AI) NR, based on experiments of the filter-bank scheme EB16 in Fig. 9, and at each channel of the module an SNR measurer and a gain reducer were incorporated. The SS gain rule from preliminary tryout was that for each channel, when $\text{SNR} > 0$ dB, the SS does nothing; when $0 \geq \text{SNR} > -3$ dB, the channel reduces gain by 4.5 dB; when $-3 \geq \text{SNR} > -6$ dB, by 7.5 dB; when $-6 \geq \text{SNR} > -12$ dB, by 13.5 dB; when $\text{SNR} \leq -12$ dB, by 18 dB. The AI strategy contains four basic intelligent activities: ① learning to master the states of incoming speech with a great probability, clear, noisy, or very noisy speech, and estimating precise SNR and noise level; ② trying out delivering the input signal into two ways: novel floating linear compression [16] for the speech-dominated ($\text{SNR} > 0$ dB) input and the multichannel SS for the noise-dominated ($\text{SNR} \leq 0$ dB) input; ③ real-time judging which way output is more quality speech, i.e., of higher SNR and more comfort loudness; ④ optimization processing to synthesize the two-way outputs with priority for next work and to feed back the priority decision to the learning process to examine correction of the previous learning perception. If this learning perception is consistent with the priority decision, e.g., speech-dominated input, the learning process qualifies for deciding the following input signal to enter one only of the two ways for a while. The multichannel SS is a key of the AI NR when the noise dominates the incoming signal.

To examine the effectiveness to suppress realistic noises, some experiments were made on the built multichannel SS module. In subsection III.A, we introduced a test speech,

which is here used for the experiments. We also acquired three test noises. The first one was from a realistic ringing wave file (Windows Ringin.wav in Windows 10 C:/Windows/Media), with an intercepting duration of 0.787 s. The second one was a milk steamer noise [13], reshaped to a middle band noise with a pink noise [17], with a duration of 0.751 s. The third one was a realistic wind noise acquired from a strong wind wave file [18], with an intercepting duration of 0.781 s. These wave files were converted into mat files and saved in our Workplace to be instantly invoked for the experiments. Due to short durations of the noises, only two words in Amy's voice, "Hi, one" were mixed up. Fig. 15 shows spectra of the noises. The black curve is the ringing spectrum of a narrow band; the blue curve is the steamer spectrum of a middle band; the brown curve is the strong wind spectrum of a wide band. Amy's voice is of SPL 60 dB, we measured its RMS (root mean square) 0.04352; then, this RMS was used to calibrate these noises SPL 60 dB. For details of how to calibrate SPL of a test signal, refer to reference [15]. For listening test by means of Adobe Soundbooth CS4, we saved the synthetical output waveforms into mat files, then converted them into wave files. For details of the files' conversion, also refer to reference [15].

Fig. 15. Spectra of three realistic noises used in SS noise reduction experiments.

A. Noise Reduction in Ringing Noise

When the speech and ringing entered the SS module together, the measured SNRs were -9.6 dB at channel 750 Hz and -1.6 dB at channel 2.25k Hz; the rest of all channels were measured $\text{SNR} > 0$ dB; according to the above SS gain rule, gain -13.5 dB was invoked at channel 750 Hz and gain -4.5 dB in 2.25k Hz; gains at the rest of all channels were not reduced to preserve the speech. Fig. 16 shows four spectra of the ringing noise and Amy's voice; since their energies mostly distribute in the low and mid frequency regions, we zoomed in the frequency axis to 5k Hz max for clearer viewing. The black curve is the input speech spectrum, and the blue is the input ringing spectrum; the brown curve is the spectrum of mixed speech and ringing at the SS input and the red is the spectrum of the mixed signals at the SS output. Comparing the two spectra of the mixed signals, the output spectrum obviously descends in the ringing energy resonance region between 500 and 1k Hz and basically recovers Amy's speech. By measuring, SNR of the SS output was 10.5 dB, which meets the conventional listening requirement. In listening test, /Hi, one/ at the output is clear to be understood.

Fig. 16. Spectra of Amy's voice and a ringing noise at SS input and output.

B. Noise Reduction in Steamer Noise

When the speech and milk steamer noise entered the SS module together, the measured SNRs at ten channels were -3.02 dB at channel 2.25k Hz, -16.7 dB at 2.75k Hz, -21.6 dB at 3.25k Hz, -20.2 dB at 3.75k Hz, -0.67 dB at 4.25k Hz, -0.5 dB at 4.75k Hz, -0.35 dB at 5.25k Hz, -0.32 dB at 5.75k Hz, -0.48 dB at 6.25k Hz, and -0.75 dB at 6.75k Hz; the other six channels had $\text{SNR} > 0$ dB. Following the SS gain rule, the ten noisy channels invoked gains -7.5 , -18 , -18 , -18 , -4.5 , -4.5 , -4.5 , -4.5 , -4.5 , and -4.5 dB, respectively; and the other six channels did nothing. Fig. 17 shows four spectra of the steamer noise and Amy's voice. We observe that the middle band noise, blue curve, has a very high lobe in a wide frequency region, 2k to 7k Hz, over the speech spectrum, black curve; the spectrum of mixed Amy's voice and steamer noise at SS input, brown curve, presents a high-energy noisy lobe but the spectrum of the mixed signals at the SS output, red curve, almost cuts away the lobe. By measuring, SNR of the multichannel SS output is 15.9 dB; by comparing, the mixed signals' spectrum at the output is much closer to Amy's voice spectrum. Such an SNR improvement in middle band noise is the best that we have seen in our researches.

Fig. 17. Spectra of Amy's voice and a milk steamer noise at SS input and output.

C. Noise Reduction in Wind Noise

When the speech and strong wind noise entered the SS module together, the measured SNRs at eight channels were -0.03 dB at channel 1.75k Hz, -16.8 dB at 2.25k Hz, -16.2 dB at 2.75k Hz, -10.2 dB at 3.25k Hz, -7.5 dB at 3.75k Hz, -2.9 dB at 4.25k Hz, -3.1 dB at 4.75k Hz, and -1.9 dB

at 5.25k Hz; the other eight channels had SNR>0 dB. Following the SS gain rule, the eight noisy channels invoked reducing gains -4.5 , -18 , -18 , -13.5 , -13.5 , -4.5 , -7.5 , and -4.5 dB, respectively; all the other channels did nothing. Fig. 18 shows four spectra of the wind noise and Amy's voice. We observe that the strong wind noise, blue curve, has higher spectrum than the speech, black curve, in the mid and high-frequency regions; then, the mixed signals' spectrum at the SS input, brown curve, is dominated by the noise in the two regions. However, after the multichannel SS processing cuts away the wide noisy spectrum, the spectrum of mixed speech and noise at the SS output, red curve, is quite closer to Amy's speech spectrum. By measuring, the multichannel SS output got SNR 6.5 dB; the remaining noise is still higher than the speech in narrow spectral regions. By listening test, we also perceived that the noise in the mixed signals at output obviously dropped down, but the output SNR is still not high enough for the conventional listening. The experimental results confirm our expectation: SNR improvement in a wide band noise may not be better than that in a middle or narrow band noise. An approach to improve this shortcoming is to utilize BW- narrower bandpass filters to more finely estimate the SNR at individual channels.

Fig. 18. Spectra of Amy's voice and a strong wind noise at SS input and output.

From the above experiments, we conclude that ① it is impossible to achieve the intelligent NR without the multichannel SS operation and the effectiveness of practical SS mostly depends on precision of SNR estimation. ② Relatively narrow BWs of the bandpass filters are beneficial to acquire details of speech and noise spectra. ③ 18 dB or more gain reduction at individual channels is not helpful to improve SNR.

V. MULTICHANNEL SCHEME FOR SPECTRAL DISTORTION REMOVAL

Everyone knows, polar patterns of a cardioid DM have a zero-gain notch at incidence 180° , independent of signal frequencies, so the DM has strong suppression of back noises. However, sensitivity (S)-gain response of the common DM at incidence 0° holds a 6 dB/octave upslope in the low and mid frequency regions as the blue curve shown in Fig 19. For comparing, S-gain response of an Omni mic is also shown in this figure, a 0 dB black horizontal line. For the S-gain definition of a mic, see reference [19]. A cross frequency of the cardioid DM and Omni mic responses is 1.78k Hz; when

frequencies <1.78 kHz, the S-gains of the DM are -18 to 0 dB. Such a sloping response must cause speech-spectrum distortion. To know the distortion severity, we made experiments. Using the blocks: Delay, Adder, Waveform Scope and Spectrum Scope, we easily built a cardioid DM test module [19]. The speech /voices/ was acquired from Amy's voice, lasting 0.641 s and sampling rate 44.1k Hz. Note that we use double slashes /a/ to represent the sound of "a". This speech has a relatively flat spectrum and is helpful to view spectral distortion. Durations of its phones /voi/, /c/, /e/, and /s/ are about 0.259, 0.109, 0.151, and 0.122 s, respectively. Fig. 20 (a) shows the waveforms of /voices/ before and after the DM, the blue and black curves, respectively; the /voice/ incidence is 0° . Comparing the two waveforms in (a), we observe that the amplitudes of /voi/ and /e/ obviously descend, but the amplitudes of /c/ and /s/ are obviously enhanced. Fig. 20 (b) shows the spectra of /voices/ after and before the DM, the black and blue curves, respectively; the black curve descends by many dB in the low and mid frequency regions and rises by about 6 dB in high-frequency region. Thus, both waveform and spectrum of the DM output suffer from the severe distortion. In addition, we listened to the sounds of two waveforms in Fig. 20 (a) separately, they were obviously different: overtone of the DM output /voices/ was perceived obviously higher than that of the input /voices/. Such a perception is consistent with the spectral change in Fig. 20 (b).

Fig. 19. S-gain responses of a cardioid DM and Omni mic.

(a) waveforms before and after DM

(b) spectra before and after DM
Fig. 20. Distortions of cardioid DM output waveform and spectrum.

(b) output spectra
Fig. 21. Outputs of the DM with an FFT plus IFFT gain balancer.

When designing a DM processor, we must solve the severe, spectral distortion; in other words, we have to balance the DM S-gain response; further, we still enable its S-gain response to match the S-gain response of an Omni mic. We designed a multichannel gain balancer, which is composed of the FFT plus IFFT scheme with frequency-domain gain table as shown in Fig.13. The FFT input of the balancer connects to the above DM module and that is our experimental module. The FFT length is 128, BWs of all the bins are 344.5 Hz, and the 64 center frequencies are 0, 344.5, 689.1, 1033.6, ... 21.7k Hz. The design key, the Frequency response table, is as follows: ① the core block has an inherent complex mode; so, the IFFT needs to accept all the FFT 128-bin outputs although the processed signals are real. ② We filled out the table so that back half of the all-bin gains of the response table are symmetrically the same as fore half of the all-bin gains. ③ According to the differences between the DM and Omni mic responses in Fig. 19, the fore 6-bin gains filled in the Frequency response table are 8.52, 4.95, 2.48, 1.67, 1.27, and 1.03 to eliminate the 6 dB/octave slope; the next 9 gains filled in the table are 0.87, 0.77, 0.69, 0.63, 0.59, 0.56, 0.53, 0.52, and 0.51 and all the rest 49 bins were filled in with the same gains 0.5 to match the Omni mic response. Fig. 21 (a) and (b) show the measured waveforms and spectra of /voices/, respectively, at outputs of this response-balanced DM and an Omni mic. We observe that in (a) the balanced waveform, blue curve, is quite closer to the Omni mic waveform, black curve; in (b) /voices/ spectrum of the balanced DM, blue curve, is quite closer to that of the Omni mic, black curve. In a word, this designed FFT plus IFFT multichannel gain balancer seamlessly removes the spectral distortion and enables the DM output to match the Omni mic output. Our listening test also confirms no perceivable difference between the input and output speeches.

(a) output waveforms

VI. CONCLUSIONS

The multichannel structures for hearing aids came up for a few decades and were developed into various schemes in applications. Nowadays, they are essential to achieve the upgraded features, such as digital NR, FB cancellation, wide DRC, and DM gain balancer. In this paper, we analyzed their principles, provided the design methods, made many experiments, examined three typical schemes, and illustrated two realistic applications by means of Simulink resource. These results acquired analytically and experimentally support the following conclusions.

1) The equal BW filter-bank structure is relatively simple, composed of an equal BW bandpass FIR filter bank, multipliers, and an adder. The key to build this structure is to design the bandpass filters. The basic design requirements include the followings: ① the frequency coverage meets the IEC and ANSI standards, e.g., 100 to 8k Hz in our scheme EB16; ② the group delays of all filters are as equal and short as possible, e.g., 5.8 ms; ③ the frequency response ripples of filter bank are as low as possible, e.g., within ± 0.8 dB; ④ the calculation load is permitted by a HW platform capacity. In this structure, number of multiplications is $N_c N_p / 2$ and number of additions is $N_c N_p / 2$ under a condition of linear phrase and direct form FIR. Our experiments confirm that the designed scheme EB16 seamlessly accomplishes the gain response assigned by a hearing aid feature. This structure can give high-level characteristics with a price of a huge design effort and heavy calculation load.

2) The Log BW filter-bank structure has the same construction as the equal BW filter-bank structure except that the filter BWs are of octave scale. This structure better matches frequency resolution of human listening and has smaller number of bank filters. On the other hand, the length of each filter is quite different and so is the group delay of each filter. Thus, this structure always suffers from delay distortion. Its design requirements are similar to those of the equal BW filter-bank structure. In our scheme LB8, it covers a frequency range of 200 to 8k Hz, the filter BWs meet 2/3 octave scale, the frequency response ripples are within ± 1.25 dB, and the filter group delays are from 0.374 to 9.25 ms. The characteristics of ripples, group delay, and sidelobe attenuation are mutually contradictory, i.e., not all can be optimized. The calculation load of this scheme is smaller than that of the equal BW scheme. The experiment of the scheme LB8 confirms the gain response assigned by a hearing aid feature. Considering the delay distortion, the selection of Log

BW filter-bank scheme needs to take more concerns.

3) The FFT plus IFFT structure needs more components: Buffer, FFT operator with Zeros-padding, Frequency-domain gain table, and IFFT operator with Selection and Add. This structure requires frame-based sample input and an IFFT synthesizer, instead of an adder, so it holds not only the equal group delays of all channels but also reconstruction characteristic. The overlap-add operation removes the temporal aliasing from FFT plus IFFT and results in the linear convolution of the input signal with frequency-domain filter, instead of a circular convolution. The heavy price that we pay for the complex structure brings back us: ① no delay distortion and zero-ripple frequency response of the synthesised output; ② smaller calculation load than that of the filter-bank structures, $N_{ft} \log_2 N_{ft} - 1.5 N_{ft} + 4$ real multiplications and $1.5 N_{ft} \log_2 N_{ft} - 1.5 N_{ft} + 2$ real additions; ③ the FFT length can be designed to be large enough and the resulting narrow BWs are helpful to obtain details of input signal spectra. The experiment of FFT plus IFFT scheme confirms the gain response assigned by a hearing-aid feature and discloses the shortcoming, inherent high sidelobe.

4) We built an experiment module of a conventional SS processor with equal BW filter-bank scheme EB16. As a key part of AI NR, this SS measures SNR at each channel and the reducing gain of the channel depends on the SNR. The preliminary SS gain rule: for each channel, when $SNR > 0$ dB, the channel does nothing; when $0 \geq SNR > -12$ dB, the reducing gain is invoked between -4.5 and -13.5 dB; when $SNR \leq -12$ dB, the reducing gain is -18 dB. Three sorts of different, realistic noises, narrow band ringing, middle band milk steamer, and wide-band strong wind, were involved to examine the effectiveness; Amy's voice mixed with them separately. The four test sounds were of SPL 60 dB. The experimental results indicate that in the ringing noise, the SNR improvement of the SS module is 10.5 dB; in the milk steamer noise, the improvement is 15.9 dB; in the wind noise, the improvement is 6.5 dB. Thus, the effectiveness of the multichannel SS processor is outstanding in the multiple noises. Before applying in a practical hearing-aid product, the further researches are needed, such as the SNR estimator, gain rule, and examination in more noises.

5) By measuring, S-gain response of a cardioid DM presents 6 dB/octave upslope from the low to mid frequencies; its output waveform and spectrum both suffer from severe, spectral distortion, which causes listening perception to deviate from the original sound. It is easy to remove this distortion by means of multichannel gain balancer. The built FFT plus IFFT gain balancer module contains the core block Frequency-Domain FIR Filter and frequency-domain gain table; the FFT length is 128 and the frequency-domain filter length is 64. Enhancing gains filled in the bins 1 to 6 of the table balance the sloping S-gain response of the cardioid DM; the reducing gains of the bins 7 to 64 ensure the DM response to match an Omni mic response. When this module combines with the DM, the response-balanced DM system forms. Our measurement confirms that waveform and spectral distortions of this DM system disappear dramatically and its full S-gain response matches Omni mic's response. Our listening test confirmed

that the system output sounds without abnormally high-frequency perception. Thus, the spectral distortion of the cardioid DM was seamlessly removed.

REFERENCES

- [1] S. A. Vorobyov, "Principles of minimum variance robust adaptive beamforming design," *Signal Processing*, vol. 93, pp. 3264-3277, 2013.
- [2] R. Rotili, E. Principi, S. Cifani, F. Piazza, and S. Squartini, *Speech Technologies*, 1st ed. INTECH Europe, Croatia, Italy, 2011, ch. 1, pp. 1-28. www.intechopen.com.
- [3] B. BabaAli, H. Sameti, and M. Safayani, "Likelihood-maximizing-based multiband spectral subtraction for robust speech recognition," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, ID 878105, pp. 1-15, 2009.
- [4] J. M. Kates and K. H. Arehart, "Multichannel dynamic-range compression using digital frequency warping," *EURASIP Journal on Applied Signal Processing*, vol. 18, pp. 3003-3014, 2005.
- [5] T. May, B. Kowalewski, and T. Dau, "Signal-to-noise-ratio-aware dynamic range compression in hearing aids," *Trends in Hearing*, vol. 22, pp. 1-12, 2018.
- [6] M. Brandgaard, "An introduction to MoreSound Intelligence™," TECH PAPER, Oticon, 2020. <https://wdh01.azureedge.net/master/whitepaper>.
- [7] M. Brandgaard, "An introduction to the Polaris™," TECH PAPER, Oticon, 2020. <https://wdh02.azureedge.net/pdf/whitepaper>.
- [8] H. Mueller and A. Gustav, "Candid round-table discuss on modern digital hearing aids and their features," *The Hearing Journal*, vol. 55, pp. 23-35, Oct. 2002.
- [9] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing - Principles, Algorithms, and Applications*, 3rd ed., PRENTICE HALL, New Jersey, 1996, ch. 6, pp. 430-460.
- [10] J. O. Smith, "Audio FFT filter banks," *Proc. of the 12th Int. Conference on Digital Audio Effects*, Como, Italy, pp. 1-4, Sept. 2009.
- [11] K. R. Rao and P. Yip, *Discrete Cosine Transform - Algorithms, Advantages, Applications*, 6th ed. ACADEMIC PRESS, San Diego, 1990, ch. 4, pp. 48-87.
- [12] Voice Samples-LAMP, Free voice for words with life, copyright by PRC-Saltito, 2020. <https://aacapps.com/lamp/voices>.
- [13] D. J. Schum, "Noise-reduction circuitry in hearing aids: (2) goals and current strategies," *The Hearing Journal*, vol. 56, pp. 32-41, Jun. 2003.
- [14] R. Bentler and Li-Kuei Chiou, "Digital noise reduction: an overview," *Trends in Amplification*, vol. 10, pp. 67-82, Jun. 2006.
- [15] Xubao Zhang, "Benefits and limitations of common directional microphones in real-world sounds," *Clinical Medicine Research*, vol. 7, pp. 103-118, Jun. 2018.
- [16] D. J. Schum and R. Sockalingam, "A new approach to nonlinear signal processing," *The Hearing Review*, vol. 17, pp. 24-32, Jul. 2010.
- [17] Audio Test Signals, Pink noise, [churchsoundcheck.com](https://www.churchsoundcheck.com/audio-test-signals.html), <https://www.churchsoundcheck.com/audio-test-signals.html>.
- [18] Free Wind Sound Effects, Mixkit-strong-wind-1156 .wav, Envato Elements, Mixkit, 2021. <https://mixkit.co/free-sound-effects/siren>
- [19] Xubao Zhang, "Spectrum distortion of a directional microphone and its removal for hearing," *Int. J. of Innovative Studies in Medical Sciences*, vol. 3, pp. 14-22, Sept. 2019.

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 as well as radar signal processing research. He worked as EA and EMC engineer with Sonova Unitron in Canada and also worked with Oticon Canada as well as Beltone Canada. Before

1993 he worked as associate professor at EE department of Xi'an Electronic Science and Technology University in radar signal processing area; before 1982 he worked as electronics engineer with Linyun Radio Manufacturer for broadcasting and navigation equipment in China. He is the author of one book and more than 40 articles.