The development of the SuperCMIT: Digitally Enhanced Shotgun Microphone with Increased Directivity

Helmut Wittek[1], Christof Faller[2], Christian Langen[1], Alexis Favrot[2], and Christophe Tournery[2]

[1] SCHOEPS Mikrofone GmbH, [2] ILLUSONIC LLC

This article is based on a lecture held at the AES convention in San Francisco, 2010 [7].

ABSTRACT

Shotgun microphones are still state-of-the-art when the goal is to achieve the highest possible directivity and S/N ratio with high signal fidelity. As opposed to beamformers, properly designed shotgun microphones do not suffer greatly from inconsistencies and sound color artifacts. A digitally enhanced shotgun microphone is proposed, using a second backward-oriented microphone capsule and digital signal processing with the goal of improving directivity and reducing diffuse gain at low and medium frequencies significantly, while leaving the sound color essentially unchanged. Furthermore, the shotgun microphone's rear lobe is attenuated in level.

1. INTRODUCTION

Shotgun microphones achieve high directivity by using a microphone capsule system with a directional response in addition to an interference tube which effectively corresponds to a continuous acoustic delay-and-sum beamformer [1]. In principle, the higher the frequency the more narrow can be the beam which is formed by the pipe. At low frequencies, where the pipe is shorter than the wave length, shotgun microphones achieve merely a supercardioid directivity pattern.

The goal of the new microphone concept is to maintain the advantages of state-of-the-art shotgun microphones, while improving the performance in those areas where a shotgun microphone has its weaknesses:

- At low frequencies, the directivity is generally no greater than that of a supercardioid.

- At low frequencies, there is a significant unwanted rear lobe (from the supercardioid-like response).

- Diffuse gain at low and mid frequencies is greater than at high frequencies.

The implementation of this concept is the new microphone SuperCMIT 2 U from SCHOEPS (referred to as SuperCMIT in the following), which was presented and released in early 2010. It will be the basis for the practical discussions and measurements in this paper.

It is described how digital signal processing is used to improve directivity at low frequencies, attenuate the rear lobe at low frequencies, and improve (decrease) diffuse gain at low and medium frequencies. The algorithm is an adapted and improved version of the one used by the highly directive two-cardioid-based microphone system previously presented in [2]. In order to achieve low delay, and to avoid time or frequency aliasing, a non-downsampled IIR filter bank is used as time-frequency representation for the processing.

The paper is organized as follows: Section 2 describes the capsule hardware and the electronic design of the microphone. The signal processing, which is applied to achieve the above-mentioned features, is described in Section 3. Evaluations and measurements are described in Section 4, and the conclusions are in Section 5.

SCHOEPS, 2011/01

2.   HARDWARE DESIGN

2.1.   Capsule elements

The SuperCMIT is a conventional shotgun microphone with an additional backward-facing cardioid and an integrated digital signal processor (DSP) which processes the signals of both microphone capsules. The two capsules are mounted nearly coincidently, *i.e.* as closely spaced as possible.

As described in [2] and specifically in Section 3, the two microphone elements are installed back-to-back. The interference tube acts on the first microphone element built in the body of the SuperCMIT. In Figure 1, the interference tube and back sound entrance are visible on the right side. (In this figure the microphone is aimed to the right.) The second microphone element is mounted directly behind the front element. This microphone element is aimed backwards (*i.e.* in the figure it is aimed to the left). As the microphone body disturbs the sound arrival at the membrane, a large front sound entrance is used to ensure free sound propagation up to medium-high frequencies. The back sound entrance for the two microphone elements is made as short as possible to minimize the distance between their membranes. The distance was reduced to less than 3 cm, which means that for frequencies below 3 kHz the geometry is sufficiently coincident.

The goal is to apply signal processing to both microphone capsules' signals at low and medium frequencies. At high frequencies only the interference tube is used. Due to this paradigm, artifacts of the adaptive algorithm are avoided and it is not necessary to further reduce the distance between the microphone elements.

2.2.   Electronic design

The electronic design of the SuperCMIT circuitry combines conventional analog condenser microphone topologies with current digital technology to enhance the shotgun microphone's signal according to the goals stated in the introduction.

The mixed-signal design is optimized for minimum clock noise interference with the analog input circuitry.

Figure 1: Two microphone elements used in the SuperCMIT.

Both microphone capsule signals are impedance converted by individual field effect transistor input circuits and led to level state-of-the-art analog-to-digital converters. The digitized capsule signals are then fed into a digital signal processor (DSP) that serves as host for the beamforming algorithm.

The DSP provides an unbalanced digital audio output that is buffered, balanced, and DC-decoupled by an AES3 transformer to provide an AES42 compliant output signal. A center tap on the secondary winding allows the microphone to be supplied with phantom power according to AES42 [3].

The supply voltage is regulated to 5 V to power the analog circuit. A highly efficient DC-DC converter generates both the DSP core supply voltage (1.25 V) as well as the supply voltage for the digital peripheral components (3.3 V).
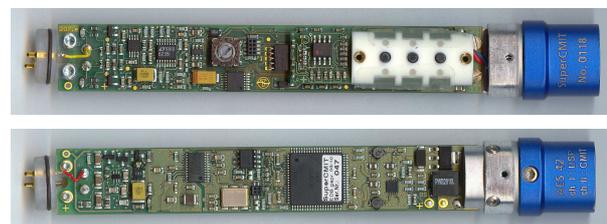

Figure 2: SuperCMIT printed circuit board, top and bottom view

The six-layer printed circuit board, shown in Figure 2 minimizes interference noise from the digital circuits into the analog circuitry. A piggyback printed circuit board featuring three buttons, visible in the top view, allows access by the user to beamforming parameters as well as filter settings.

3. SIGNAL PROCESSING

The signal processing used to enhance the shotgun signal is described in the following. The shotgun signal is denoted $f(n)$ (forward-facing microphone signal) and the cardioid signal is denoted $b(n)$ (backward-facing microphone signal). Time-frequency processing is used to simulate, based on $f(n)$ and $b(n)$, a highly directive microphone with controllable directivity and diffuse response. In [2] the implementation of this principle is described based on a short-time Fourier transform. For the SuperCMIT the delay requirement was so strict that a short-time Fourier transform was not suitable for the task.

3.1. Filter Bank

An IIR filter bank was developed without using any downsampling, based on techniques described in [4,5,6]. The filter bank yields nine subbands. The top panel in Figure 3 shows the magnitude/frequency response of the subbands and the all-pass response of the sum of all subbands (bold). The group delay response of the unmodified filter bank output signal (sum of all subbands) is shown in the bottom panel of the figure.

For the signal $f(n)$ the corresponding subband signals are denoted $F_i(n)$, where subband index $i=0$ corresponds to the lowest-frequency subband and $i=8$ corresponds to the highest-frequency subband. The subbands are similarly defined for the signal $b(n)$. The subbands of the filter bank are "doubly complementary" [5], *i.e.*

$$\left| \sum_{i=0}^{8} S_i(z) \right| = 1 \quad \text{and} \quad \sum_{i=0}^{8} \left| S_i(z) \right|^2 = 1 \quad , \ (1)$$

where $S_i(z)$ is the *z*-transform of the impulse response of subband IIR filter i. The first property in (1) ensures all-pass behavior of the synthesis (sum) output signal, and the second property implies frequency-separating subbands (the data in the top panel of Figure 3 implies both properties).
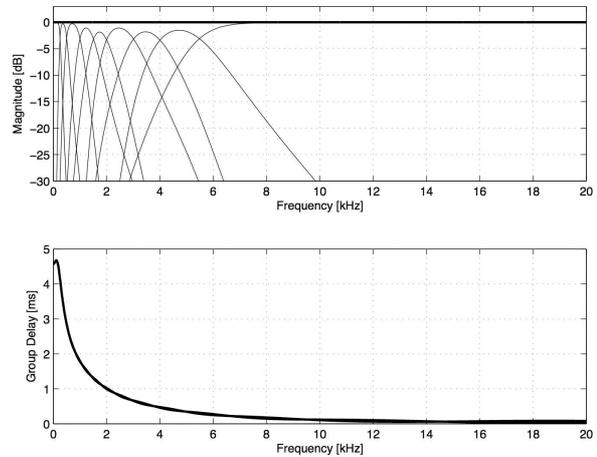

Figure 3: Magnitude response of the subbands (thin, top), filter bank output (bold, top), and group delay (bottom) are shown

3.2. Directivity enhancement and diffuse attenuation processing

The goal is to improve directivity, attenuate the rear lobe, and decrease the diffuse gain at low and mid frequencies, where shotgun microphones do not perform as well as at high frequencies. Processing is applied in the eight lower subbands, whereas the ninth high-frequency subband is not processed, since the previously mentioned weaknesses of shotgun microphones appear only at lower frequencies.
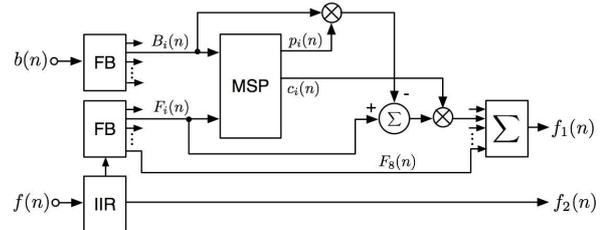

Figure 4: Schematic diagram of the processing that is applied to the forward and backwards facing signals $f(n)$ and $b(n)$ to generate the output signals. FB, IIR, and MSP denote filterbank, IIR filters, and microphone signal processing, respectively

The technique described in [2] uses two cardioid microphone signals, facing forward and backward, to generate a "virtual microphone" signal facing forward, which is highly directive and has controllable diffuse response. The technique was specifically

adapted for the SuperCMIT. Figure 4 illustrates the processing that is applied to the signals *f(n)* and *b(n)*. First, various IIR filters (optional low cut and high shelving filters, etc.) are applied to the *f(n)* signal. Then, the previously described IIR filter banks are applied, resulting in the subband signals $F_i(n)$ and $B_i(n)$. The eight lower subbands, $0 \leq i < 8$ are processed as follows:

- A predictor is used to predict $F_i(n)$ from $B_i(n)$. The predictor's magnitude is limited to achieve the desired directivity and rear-lobe attenuation (see [2] for more details). The limited predictor is denoted $p_i(n)$.

- The limited predictor $p_i(n)$ is applied to the signal $B_i(n)$ and the resulting signal is subtracted from the forward-facing signal $F_i(n)$:

$$F_i(n) - p_i(n)B_i(n) \ .$$

- A post-scaling factor $c_i(n)$ is computed to achieve the desired diffuse response. The post-scaling factor is applied to the previously computed signal:

$$F_{1,i}(n) = c_i(n)\big(F_i(n) - p_i(n)B_i(n)\big) \ .$$

The processed subbands $F_{1,i}(n)$ are summed and added to the non-processed high frequency subband $F_8(n)$ to generate the enhanced shotgun output signal $f_1(n)$. The second output channel $f_2(n)$ is the shotgun input signal *f(n)* which has merely been IIR-filtered.

Note that the parameters of the processing (such as directivity and diffuse-field response) are chosen in each subband individually to achieve similar processed results despite the variations in the two microphone channels' frequency characteristics.

4. EVALUATIONS AND MEASUREMENTS

The following section will present frequency response curves and polar diagrams of the SuperCMIT microphone and will also discuss the subjective evaluation. In this microphone, the two output channels contain both the pure shotgun signal (on channel 2) as well as the output of the beamforming algorithm (on channel 1). Two different settings can be chosen for the beamforming algorithm. Preset 1 is a moderate setting, with moderately enhanced directivity and a moderate decrease of the diffuse-field level. Preset 2 is the "strong" setting, with clearly enhanced directivity and a large decrease of the diffuse-field level.

The measurements of the free-field frequency response clearly show the behavior expected from the theoretical models. The top diagram in Figure 5 shows the frequency response curves of the unprocessed shotgun microphone channel. The black curve indicates the 0° free-field frequency response with a constant sensitivity up to the frequency range of 5 kHz. The high-frequency boost up to 15 kHz is significant for shotgun microphones, both to compensate for high-frequency losses caused by windscreens and to avoid subjective 'dullness' of sound that would otherwise be caused by the increasing directivity at high frequencies.

The red curve shows the 90° free-field frequency response with a significant high-frequency rolloff. Small irregularities in the response are caused by the interference tube, which produces increasing directivity at high frequencies. At a sound incidence angle of 180° (green curve) the sensitivity is increased in comparison to the 90° response; at low frequencies the directivity tends toward a supercardioid polar pattern.

The beamforming algorithm does not change the frequency response at 0°, as can be seen in the middle diagram. The frequency response is identical to the response of the shotgun shown in the upper graph. On the other hand, the 90° free-field frequency sensitivity is reduced to the level of the frequency range above 6 kHz where the shotgun principle works without compromise—the 'seamless attachment' of the beamformed signal to the shotgun microphone's signal shows a nearly flat frequency response within ±3 dB. The 180° free-field frequency response has the same 'characteristic' as the shotgun's response without beamforming, but is reduced in sensitivity by 4 dB. These characteristics lead to unchanged sonic behavior of the SuperCMIT as compared to a classic shotgun microphone, but with significantly higher directivity at low frequencies.

The bottom diagram of Figure 5 shows the free-field frequency response curves produced by Preset 2, the "strong" setting of the beamforming algorithm. Again the 0° free-field response is identical to the pure interference tube microphone signal, while the 90° sensitivity is reduced by another 4 dB.
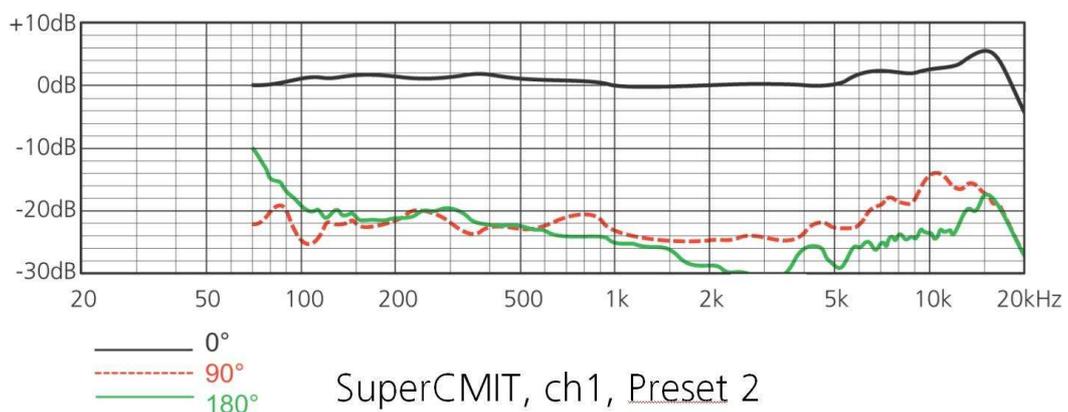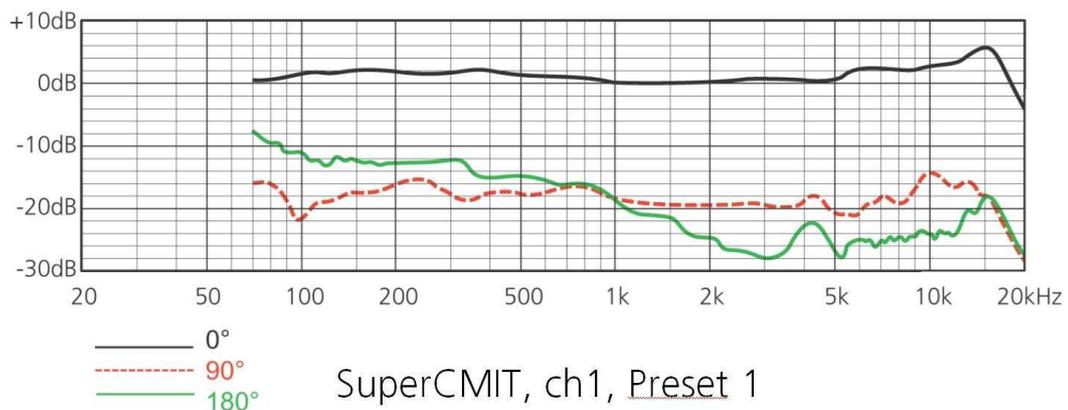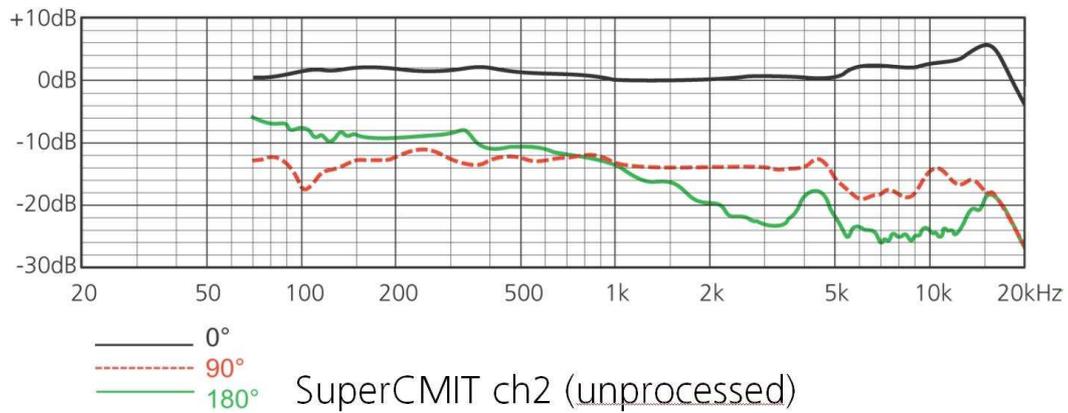
Figure 5: Free-field frequency responses of the SuperCMIT at three sound incidence angles (black—0°, red—90°, green—180°):

Top diagram: SuperCMIT channel 2 (unprocessed shotgun)
Middle diagram: SuperCMIT channel 1, Preset 1 (moderately enhanced directivity)
Bottom diagram: SuperCMIT channel 1, Preset 2 (strongly enhanced directivity).

SCHOEPS, 2011/01

Even if this does not seem very impressive at first glance, the great benefit of the enhanced beamforming algorithm lies in the reduction of 180° sensitivity by up to 10 dB at low frequencies. Because the high-frequency directivity is even greater than that of the interference tube, this setting is suitable only for applications that require an extremely high suppression of ambient noise. The sensitivity reduction of 20 dB for 90° and 180° sound incidence in the entire frequency range up to 8 kHz results to a polar pattern similar to the front lobe of a figure-eight microphone characteristic (see **Figure 6**).

Since the polar patterns of the beamformed channel match closely with those of the unprocessed interference tube, but with significantly higher directivity (as can be seen in **Figure 6**), the enhanced beamformer almost completely removes the back lobe of the microphone's polar pattern.
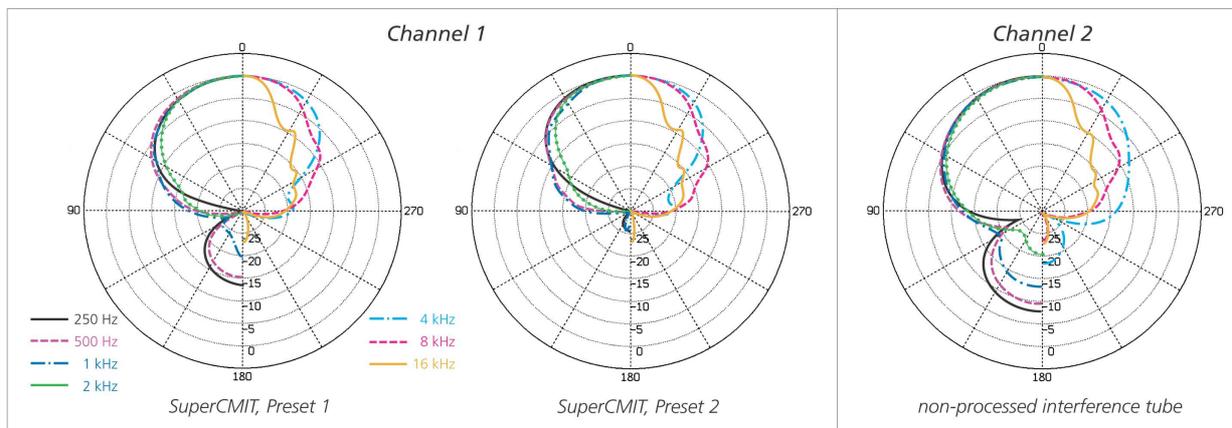

Figure 6: **Polar patterns of SuperCMIT beamforming channel (left diagram), enhanced beamformer channel (right diagram).**
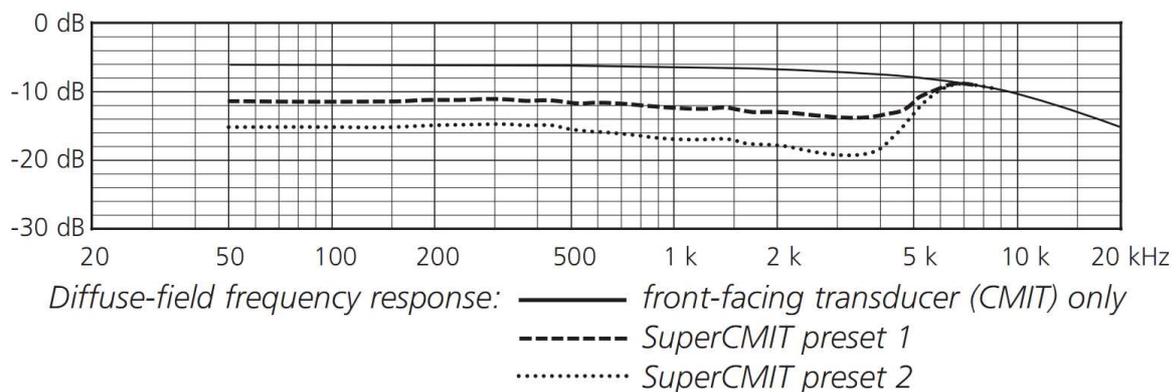

Figure 7: Diffuse-field frequency responses of SuperCMIT ch2, SuperCMIT ch1 Preset 1, SuperCMIT ch1 Preset 2.

In practice the reduction of the diffuse sound level proves to be even more important than the polar pattern. This seems to call for a new understanding of the issues involved. For conventional microphone patterns, there exists a direct relationship between the polar pattern and the diffuse sound level. This is different in the case of the SuperCMIT, because the diffuse sound level can be reduced by a higher degree. Figure 7 shows the diffuse sound frequency responses of the three possible SuperCMIT output signals.

This performance is remarkable when comparing it to conventional first-order microphones. Figure 8
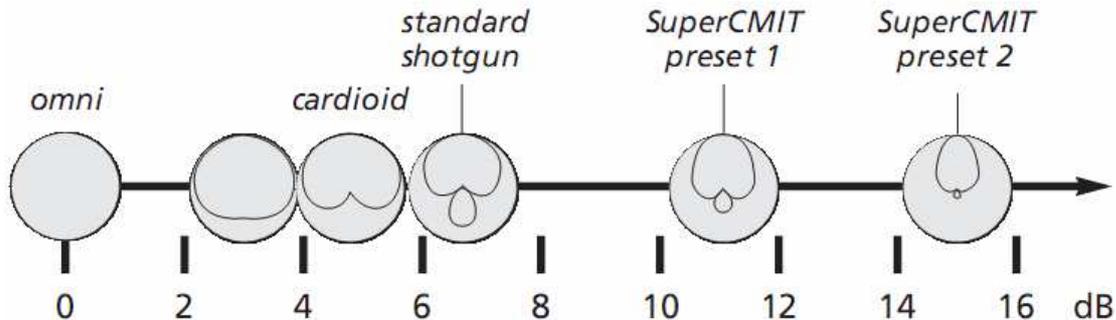
Figure 8: **Free-to-diffuse field ratio for first-order directivities and SuperCMIT beamformers**

shows the directivity index, *i.e.* the overall free-to-diffuse field ratio, for omnidirectional to supercardioid polar patterns as well as conventional shotgun microphones and the SuperCMIT.

The highest ratio possible for first-order directional microphones is a reduction of the diffuse sound by 6 dB. Even interference-tube microphones cannot improve this ratio significantly, since their principle works only at frequencies high enough for their wavelengths to be within the range of the interference tube's length.

The beamforming algorithm helps to increase that ratio to 11 dB, since the received diffuse sound energy is reduced by a factor of 12.6. This ratio is further increased by the enhanced beamforming algorithm, leading to a ratio of as much as 15 dB—a factor of 31.6.

Subjective evaluations

The choice of parameters and the definition of the presets were the result of elaborate recording and listening sessions, including a selected panel of distinguished expert listeners. The timbral fidelity of the processed signal could be ensured by a suitable mechanical design and the design of the algorithm. Temporal artifacts of the processing can be made audible only when very radical parameter settings are chosen. This is successfully avoided with Preset 1, which proved to be solid without any audible artifacts. In spite of possible audible artifacts, Preset 2 enables a further increase in directivity and decrease of diffuse-field level. Therefore it is a setting for special applications—only for recording situations which demand such strong settings, where possible arti-

facts will be overlooked or masked by other audio signals. Furthermore, sometimes a sound field with too little reverberation can sound unnatural, in particular in small rooms.

Recent experience suggests that for film applications Preset 1 is chosen most of the time, whereas for sports and ENG applications, Preset 2 is also used. Since the algorithm features flexible parameters for controlling directivity, diffuse response, and rear-lobe attenuation, future firmware might realize further settings.

5.   CONCLUSIONS

A new microphone principle, with the goal of enhancing a shotgun microphone by means of a second microphone element and digital signal processing, was presented. At low and medium frequencies, where a shotgun microphone has its weaknesses, signal processing is used to improve directivity and direct-to-diffuse gain, and to attenuate the rear lobe.

Both technical data and subjective evaluations have shown that a feasible implementation of this principle is possible. Initial feedback from users confirms the usefulness and applicability of the increased directivity in standard recording situations.

Future work will further study the circumstances under which some temporal artifacts exist when the strong beamformer setting is used (Preset 2). This would enable more predictable and controlled application of Preset 2 of the SuperCMIT microphone.

REFERENCES

[1] K. Tamm and G. Kurtze, "Ein neuartiges Mikrophon großer Richtungsselektivität," Acustica, No. 5, Vol. 4, 469-470, Beiheft 1, 1954.

[2] C. Faller, "A highly directive 2-capsule based microphone system," in Preprint 123rd Conv. Aud. Eng. Soc., Oct. 2007.

[3] AES 42 Standard: http://www.aes.org/publications/standards/search.cfm?docID=38

[4] Y. Neuvo and S. K. Mitra, "Complementary iir digital filters," in Proc. IEEE Int. Symp. on Circuits and Systems, May 1984, pp. 234–237.

[5] P. A. Regalia, S. K. Mitra, P. P. Vaidyanathan, M. K. Renfors, and Y. Neuvo, "Tree-structured complementary filter banks using all-pass sections," IEEE Trans. Circuits Syst., vol.34, pp. 1470–1484, December 1987.

[6] C. Favrot and C. Faller, "Complementary N-band IIR filterbank based on 2-band complementary filters," in Proc. Intl. Works. on Acoust. Echo and Noise Control (IWAENC), Aug. 2010.

[7] H. Wittek , C. Faller, A. Favrot, C. Tournery, C. Langen, "Digitally Enhanced Shotgun Microphone with Increased Directivity" in Preprint 129th Conv. Aud. Eng. Soc., Nov. 2010.

SCHOEPS, 2011/01