DISTRIBUTED SPATIAL AUDIO CODING IN WIRELESS HEARING AIDS

Olivier Roy [†] *and Martin Vetterli* [†§]

[†] School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

[§] Department of Electrical Engineering and Computer Sciences
University of California at Berkeley (UCB), Berkeley, CA 94720, USA

{olivier.roy, martin.vetterli}@epfl.ch

ABSTRACT

The information content of binaural signals can be beneficial to many algorithms deployed in current digital hearing aids. However, the exchange of such signals over a wireless communication link requires transmission schemes that must fulfill demanding technical constraints. We present a distributed coding algorithm that builds on psychoacoustic principles in order to achieve this goal with low bitrates, while preserving affordable complexity. The key steps of the proposed algorithm are detailed and the accuracy of the signal exchange mechanism is evaluated in simple simulated acoustic scenarios.

1. INTRODUCTION

Most current hearing aid systems involve signal processing algorithms that run independently at each hearing device. However, the information gained from comparing signals recorded at both ears of the user could potential allow significant improvements over existing solutions. For example, the detection of spurious frequencies arising from acoustic feedback may be improved by analyzing binaural signals [1]. Also, the interference mitigation of noise reduction algorithms could be enhanced by combining signals from the left and right hearing aid [2]. In this context, the use of wireless technology to connect the two hearing instruments offers new perspectives to many of the challenging problems encountered in practice.

Motivated by the above considerations, we present a practical coding scheme that enables two hearing aids to exchange their recorded signal by means of a rate-constrained communication link. The ultimate goal of the algorithm is to provide both devices with reconstructed binaural signals that are perceptually equivalent to the original ones. In addition, the method should tradeoff bitrate and processing delay while keeping complexity at an acceptable level. To this end, our method builds upon the psychoacoustic fundamentals exposed in [3]. We show that, for simple acoustic scenarios, one hearing aid can recover the signal available at the other device by imposing appropriate spatial cues on its own signal. The communication protocol hence amounts to exchange the information needed to properly estimate these cues. Moreover, the spatial correlation induced by the hearing aid setup on the recorded signals can be further exploited to reduce the transmission bitrate, as demonstrated by seminal work on distributed source coding [4, 5]. In particular, we explain how the shadowing effect of the head can be taken into consideration to decrease the communication rate. It should be noted that general distributed coding schemes have already been proposed (see e.g. [6]) but their wide applicability usually comes at the expense of a high latency which make their deployment difficult in real-time systems. The effectiveness of the proposed approach is finally assessed through simulations using acoustic scenes synthesized from speech excerpts and head-related transfer functions (HRTF).

This research was supported by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS, http://www.mics.org), a center supported by the Swiss National Science Foundation under grant number 5005-67322.

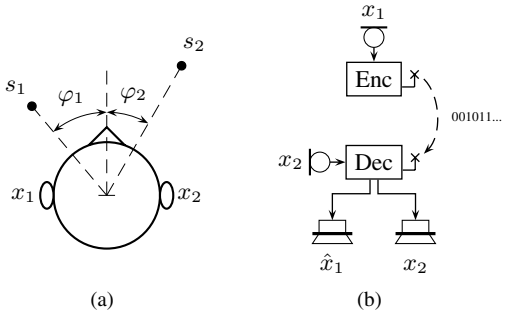


Figure 1: Signal exchange between wireless hearing aids. (a) Typical recording setup (here with $M = 2$ sources). (b) Signal encoding from the perspective of one hearing aid 1.

The paper is organized as follows. In Section 2, the signal exchange problem is precisely stated. Section 3 describes the main parts of the proposed algorithm and provides implementation details. Simulation results are then presented in Section 4. Conclusions are drawn in Section 5 and future directions of research are briefly outlined.

2. THE SIGNAL EXCHANGE PROBLEM

The general setup of interest is illustrated in Figure 1(a). We consider an acoustic scene composed of multiple sound sources $s_1, s_2, \ldots, s_M$ and denote by x_1 and x_2 the signals recorded respectively at the user's left and right hearing aid (hereafter referred to as hearing aid 1 and 2). The two instruments wish to exchange their signal by means of a wireless communication link. Owing to the inherent symmetry of the problem, the rest of the paper will adopt the perspective of one hearing device (say hearing aid 1) and will develop a coding method that allows to efficiently transmit x_1

978-1-4244-1619-6/07/$25.00 ©2007 IEEE

provided that x_2 is available at the decoder, that is at hearing aid 2. In this case, the coding setup reduces to that depicted in Figure 1(b). Hearing aid 1 maps x_1 into a bit stream, which is then wirelessly transmitted to hearing aid 2. Based on the received data and its own signal x_2, this latter reconstructs the signal $\hat{x}_1$. Perceptual experiments allow to assess the reconstruction accuracy by comparing the original binaural signals (x_1, x_2) to that obtained at hearing aid 2, namely $(\hat{x}_1, x_2)$. The next section presents the details of the proposed algorithm.

3. DISTRIBUTED SPATIAL AUDIO CODING

3.1. Overview

It has been shown in [3] that the perceptual spatial correlation between x_1 and x_2 can be well captured by cues referred to as *inter-channel level difference* (ICLD) and *inter-channel time difference* (ICTD). If an encoder has access to both x_1 and x_2, the strategy adopted in [7] consists in sending those cues along with the sum (mono) signal $x_1 + x_2$. The original signals can then be recovered at the decoder by imposing those cues on the sum signal. A significant bitrate saving can be achieved by realizing that ICLDs and ICTDs vary slowly across time and frequency and thus only need to be estimated on a time-frequency atom basis. The strategy adopted in our case is similar except for two major differences: (i) the cues must be estimated in a distributed fashion since x_1 and x_2 are not available centrally and (ii) x_1 is recovered by applying the spatial cues on the signal x_2 available at the decoder.

3.2. Time-Frequency Processing

The processing in the proposed algorithm is performed on a time-frequency representation obtained as follows. First, the length N discrete-time input signals x_1 and x_2 are segmented into overlapping frames of even length N_f by applying a window

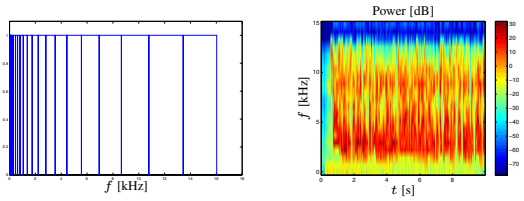
$$w[k] = \begin{cases} \sin^2\left(\frac{(k-N_z)\pi}{N_w}\right), & N_z \le k < N_f - N_z, \\ 0, & 0 \le k < N_z \text{ or} \\ & N_f - N_z \le k < N_f. \end{cases}$$

This corresponds to a Hann window of even length N_w with N_z zeros added on each side to enable delay synthesis in the discrete Fourier transform (DFT) domain. Here $N_f = N_w + 2N_z$. Consecutive frames are obtained by shifting the analysis window by $N_w/2$ samples (50% overlap). This allows perfect reconstruction of the input signal in an overlap-add framework since the shifted windows sum up to 1. Finally, a N_f-point DFT is taken on each frame which results in the time-frequency representations $X_1[n,k]$ and $X_2[n,k]$ for time index $n = 0, 1, \ldots, \lfloor (2N/N_w) \rfloor - 1$ and frequency index $k = 0, 1, \ldots, N_f - 1$. Since the input signals are real-valued, the spectrum is symmetric and only the first $N_f/2+1$ frequency coefficients need to be considered.

3.3. Analysis

The multichannel audio coding scheme presented in [7] demonstrates that estimating a single spatial cue for a group of adjacent frequencies is sufficient for a perceptually transparent reconstruction. At each frame n, the $N_f/2 + 1$ frequency indexes are grouped in frequency subbands according to a partition $\mathcal{B}_l$ $(l = 0, 1, \ldots, L-1)$, i.e. such that

$$\bigcup_{l=0}^{L-1} \mathcal{B}_l = \{0, 1, \ldots, N_f/2\} \quad \text{and} \quad \mathcal{B}_l \cap \mathcal{B}_{l'} = \emptyset \text{ for all } l \neq l'.$$



(a) (b)

Figure 2: Time-frequency processing. (a) Partitioning of the frequency band in critical subbands. (b) Power estimates as a function of time and frequency.

Note that, in the sequel, frequency subbands are always indexed with l whereas frequency bins are indexed with k. Psychoacoustic experiments suggest that spatial perception is most likely based on a frequency subband representation with bandwidths proportional to the the critical bandwidth of the auditory system [8]. Since this latter can be approximated by the equivalent rectangular bandwidth (ERB) [9], we use a constant bandwidth of N_b [ERB] to obtain a non-uniform partitioning of the frequency band according to the relation

$$N_b(f) = 21.4 \log_{10}(0.00437 f + 1),$$

where f is the frequency measured in Hertz. This is shown in Figure 2(a). The analysis part of our algorithm at time n simply consists in computing at both hearing aids an estimate of the signal power, in dB, for each critical band $\mathcal{B}_l$ as

$$p_1[n,l] = 10 \log_{10}\left(\frac{1}{|\mathcal{B}_l|}\sum_{k\in\mathcal{B}_l} |X_1[n,k]|^2\right) \quad \text{and}$$

$$p_2[n,l] = 10 \log_{10}\left(\frac{1}{|\mathcal{B}_l|}\sum_{k\in\mathcal{B}_l} |X_2[n,k]|^2\right).$$

A typical representation of such power estimates is depicted in Figure 2(b). Note that $p_1[n,l]$ and $p_2[n,l]$ will allow to compute ICLDs at hearing aid 2 for each critical band. As explained later, the ICTDs will be inferred from ICLDs using a simple HRTF lookup table. This strategy requires no additional information to be sent. The communication bitrate is hence reduced to a bare minimum.

3.4. Coding

We now explain how hearing aid 1 can efficiently encode its power estimates at time n taking into account the specificities of the hearing aid recording setup. These power estimates will be necessary for the computation of ICLDs at hearing aid 2. The key is to observe that, while $p_1[n,l]$ and $p_2[n,l]$ may vary significantly as a function of the critical band index l, the ICLDs, defined as

$$\Delta p[n,l] = p_1[n,l] - p_2[n,l],$$

are bounded above (resp. below) by the level difference caused by the head when a source is on the far left (resp. the far right) of the user. Let us denote by $h_{1,\varphi}[n]$ and $h_{2,\varphi}[n]$ the left and right head-related impulse responses (HRIR) at elevation zero and azimuth φ, and by $H_{1,\varphi}[k]$ and $H_{2,\varphi}[k]$ the corresponding HRTFs. The ICLD in critical band l can be computed as a function of φ as

$$\Delta p_\varphi[l] = 10\log_{10}\frac{\frac{1}{|\mathcal{B}_l|}\sum_{k\in\mathcal{B}_l}|H_{1,\varphi}[k]|^2}{\frac{1}{|\mathcal{B}_l|}\sum_{k\in\mathcal{B}_l}|H_{2,\varphi}[k]|^2} \quad (1)$$

and is thus contained in the interval given by[1]

$$\left[\,\Delta p_{min}[l]\,,\Delta p_{max}[l]\,\right] = \left[\,\Delta p_{90}[l]\,,\Delta p_{-90}[l]\,\right]. \quad (2)$$

In the centralized scenario, ICLDs can hence be quantized by a uniform scalar quantizer with range (2).

In our case, an equivalent bitrate saving can be achieved using a modulo approach. The powers $p_1[n,l]$ and $p_2[n,l]$ are quantized using a uniform scalar quantizer with range $[p_{min}, p_{max}]$ and stepsize s. The range can be chosen arbitrarily but must be large enough to accommodate all relevant powers. The resulting quantization indexes $i_1[n,l]$ and $i_2[n,l]$ satisfy

$$\begin{aligned} i_1[n,l] - i_2[n,l] &\in \left\{\Delta i_{min}[l], \Delta i_{max}[l]\right\} \\ &= \left\{\left\lfloor\frac{\Delta p_{min}[l]}{s}\right\rfloor, \left\lceil\frac{\Delta p_{max}[l]}{s}\right\rceil\right\}. \end{aligned} \quad (3)$$

Since $i_2[n,l]$ is available at the decoder, hearing aid 1 only needs to transmit a number of bits that allow hearing aid 2 to choose the correct index among the set of candidates whose cardinality is given by

$$\bar{\Delta} i[l] = \Delta i_{max}[l] - \Delta i_{min}[l] + 1\,.$$

This can be achieved by sending the value of the indexes $i_1[n,l]$ modulo $\bar{\Delta} i[l]$, i.e. using only $\log_2 \bar{\Delta} i[l]$ bits. This strategy thus permits a bitrate saving equal to that of the centralized scenario. Moreover, at low frequencies, the shadowing effect of the head is less important than at high frequencies. The corresponding $\bar{\Delta} i[l]$ is thus smaller and the number of required bits can be reduced. Therefore, the proposed scheme takes full benefit of the characteristics of the recording setup. From an implementation point-of-view, a single scalar quantizer with stepsize s is used for all critical bands. The modulo strategy simply corresponds to an index reuse, as illustrated in Figure 3. At the decoder, the index $i_2[n,l]$ is first computed and among all possible indexes $i_1[n,l]$ satisfying relation (3), the one with the correct modulo is selected. The reconstructed power estimates are denoted $\hat{p}_1[n,l]$.

3.5. Synthesis

The synthesis part of the algorithm aims at recovering, at hearing aid 2, the time-frequency spectrum $\hat{X}_1[n,k]$ using $X_2[n,k]$ and the reconstructed power estimates $\hat{p}_1[n,l]$. This is achieved as follows. First, ICLDs are computed in critical bands as

$$\Delta\hat{p}[n,l] = \hat{p}_1[n,l] - p_2[n,l]\,, \quad (4)$$

[1] The azimuths are measured in the trigonometric direction and the zero angle corresponds to the front. Note that 90° is chosen for simplicity but can be replaced by any other angle at which the level difference is maximum in absolute value.

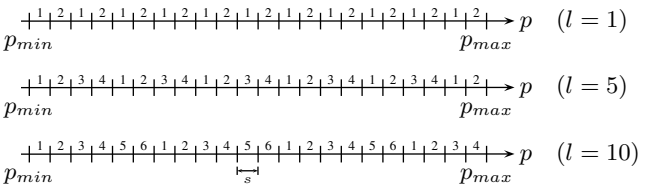


Figure 3: Illustration of the proposed modulo coding approach. The power p is always quantized using a scalar quantizer with range $[p_{min}, p_{max}]$ and stepsize s. Indexes, however, are assigned modulo the ICLD range $\bar{\Delta} i[l]$ specific to each critical band. In this example, the index reuse for $l = 1$ (low frequencies) is more frequent than at $l = 10$ (high frequencies).

for $l = 0, 1, \ldots, L-1$. Suitable interpolation is then applied to obtain the ICLDs $\Delta\hat{p}[n,k]$ over the entire frequency band, i.e. for $k = 0, 1, \ldots, N_f/2$. These ICLDs are then synthesized as

$$\hat{X}_{1a}[n,k] = X_2[n,k]10^{\frac{\Delta\hat{p}[n,k]}{20}}\,. \quad (5)$$

In order to provide an accurate rendering of the acoustic scene in real scenarios, ICLDs are not sufficient. Phase differences between the two signals must also be synthesized. To this end, we resort to an HRTF lookup table that allows to map the computed ICLDs to ICTDs. For each critical band l, we first compute the ICLD given by (1) for a virtual source at different azimuths $\varphi \in \mathcal{A}$. We then select the ICLD closest to that obtained in (4). The chosen azimuthal angle, denoted $\hat{\varphi}_l$, hence follows from

$$\hat{\varphi}_l = \arg\min_{\varphi\in\mathcal{A}} |\Delta\hat{p}[n,l] - \Delta p_\varphi[l]|\,.$$

The corresponding ICTD, denoted $\Delta\hat{\tau}_a[n,l]$ and expressed in samples, is then computed as the difference between the positions of the maxima in the HRIRs of this virtual source, namely

$$\Delta\hat{\tau}_a[n,l] = \arg\max_m |h_{1,\hat{\varphi}_l}[m]| - \arg\max_m |h_{2,\hat{\varphi}_l}[m]|\,.$$

Note that the above operations can be implemented by means of a simple lookup table where the relevant ICLD-ICTD pairs are pre-computed for the set of azimuths $\mathcal{A}$. Similarly to the ICLDs, ICTDs $\Delta\hat{\tau}_a[n,k]$ are obtained for all frequencies by interpolation. The ICTDs are then applied to the spectrum obtained in (5) as

$$\hat{X}_{1b}[n,k] = \hat{X}_{1a}[n,k]e^{-j\frac{2\pi}{N_f}k\Delta\hat{\tau}_a[n,k]}\,. \quad (6)$$

However, in order to have smoother variations over time and to take into account the power of the signals for time-delay synthesis, the spectrum given by equation (6) is further processed. More precisely, we treat it as the true spectrum and compute a smoothed estimate of the cross power spectral density S_{12} between x_1 and x_2, namely

$$S_{12}[n,k] = \alpha\hat{X}_{1b}[n,k]X_2^*[n,k] + (1-\alpha)S_{12}[n-1,k]\,,$$

where the superscript * denotes the complex conjugate. At initialization, $S_{12}[0,k]$ is set to zero for all k. Let us denote by $\measuredangle S_{12}[n,k]$ the phases of S_{12}. The final ICTDs $\Delta\hat{\tau}[n,l]$ are obtained by grouping the phases in critical bands and perform a least mean-squared fitting through zero for each band. The slopes of the fitted lines correspond to the ICTDs. We obtain

$$\Delta\hat{\tau}[n,l] = \frac{N_f}{2\pi}\frac{\sum_{k\in\mathcal{B}_l} k\measuredangle S_{12}[n,k]}{\sum_{k\in\mathcal{B}_l} k^2}\,.$$

Since ICTDs are most important at low frequencies [8], we only synthesize them up to a maximum frequency f_m. For sufficiently small f_m, the phase ambiguity problem can thus be neglected. Finally, the interpolated values $\Delta\hat{\tau}[n,k]$ allow to reconstruct the spectrum from equation (5) as

$$\hat{X}_1[n,k] = \hat{X}_{1a}[n,k]e^{-j\frac{2\pi}{N_f}k\Delta\hat{\tau}[n,k]}$$

and the time-domain output signal $\hat{x}_1[n]$ can be computed.

3.6. Discussion

It is important to point out that the distributed coding scheme described above assumes that the ICLDs belong to the range given by (2). If this condition is violated the wrong index will be reconstructed, possibly leading to audible artifacts in the output signals. This may happen, for example, if the ICLD range is chosen too small. The choice of suitable values for the range across critical bands hence provides a way to tradeoff the aggressiveness of the coding scheme with the reconstruction accuracy. In particular, we may benefit from the interactive nature of the communication link established between the two hearing aids to adapt the coding scheme dynamically. We might also envision a combined use of standard and distributed coding methods.

It should also be noted that, while the HRTF table lookup strategy described above seems satisfactory for simple acoustic scene rendering, coding of the true ICTDs are required for more realistic scenarios (e.g. with reverberation). These issues are however matters of current research.

4. SIMULATION RESULTS

In order to assess the reconstruction quality of the proposed signal exchange algorithm, we performed some experiments with binaural signals synthesized from speech excerpts and HRTFs obtained from the database in [10]. The parameters of Section 3 are chosen as follows. The sampling rate is $f_s = 32$ [kHz], the frame size is $N_f = 1024$ samples with a window size of $N_w = 896$ samples. The induced algorithmic delay (i.e. excluding the time needed for computations and wireless transmission) is thus $N_w/f_s = 28$ [ms]. The partition bandwidth is set to $N_b = 2$ [ERB] which corresponds to $L = 21$ critical bands spanning frequencies up to 16 [kHz]. The HRTF lookup table maps ICLDs to ICTDs for 72 uniformly spaced azimuths on the horizontal plane (elevation zero). Finally, ICTD synthesis is applied up to $f_m = 1.5$ [kHz] and the cross power spectral density smoothing factor is set to $\alpha = 0.3$. Regarding the distributed coding scheme, the ICLDs are assumed to take values in intervals of linearly increasing length as a function of the critical band index l, namely from $[-5, 5]$ [dB] at $l = 1$ to $[-35, 35]$ [dB] at $l = L = 21$. The quantizer stepsize s is chosen such as to meet a desired bitrate R, here set to $R = 8$ [kb/s].

Three simulations were performed with 1, 2 and 3 speech sources at $\varphi = 30^\circ$, $(\varphi_1, \varphi_2) = (-30^\circ, 30^\circ)$ and $(\varphi_1, \varphi_2, \varphi_3) = (-30^\circ, 0^\circ, 30^\circ)$, respectively. Informal listening indicates that the proposed algorithm renders the binaural signals with a similar spatial image and only few artifacts. We noted however that the spatial width of the synthesized auditory scene tends to be slightly more narrow than the original one, in particular when multiple sources are present. This may be explained by the fact that when two oppositely located sources are concurrently active in a time-frequency atom, the corresponding ICLDs (hence ICTDs) tend to average each other out. A possible solution may consist in designing a selection mechanism to synthesize ICTD cues only when they actually correspond to a physical sound source. This will however be addressed in future work.

5. CONCLUSIONS

We presented an algorithm that allows the exchange of acoustic signals between two hearing aids linked with a wireless communication medium. The proposed scheme capitalizes on the spatial correlation between the recorded signals in order to reduce the transmission bitrate. In order to improve the spatial rendering of the method in more complex scenarios, current research is focussing on the development of distributed coding schemes for ICTDs.

6. ACKNOWLEDGMENTS

We would like to thank Christof Faller for enlightening discussions.

7. REFERENCES

[1] V. Hamacher, J. Chalupper, J. Eggers, E. Fischer, U. Kornagel, H. Puder, and U. Rass, "Signal processing in high-end hearing aids: State of the art, challenges, and future trends," *EURASIP Journal on Applied Signal Processing*, vol. 18, pp. 2915–2929, 2005.

[2] O. Roy and M. Vetterli, "Rate-constrained beamforming for collaborating hearing aids," *IEEE International Symposium on Information Theory*, pp. 2809–2813, July 2006.

[3] F. Baumgarte and C. Faller, "Binaural cue coding - Part I: Psychoacoustic fundamentals and design principles," *IEEE Trans. Speech Audio Processing*, vol. 11, no. 6, pp. 509–519, November 2003.

[4] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, pp. 471–480, July 1973.

[5] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, January 1976.

[6] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUSS): design and construction," *IEEE Trans. Inform. Theory*, vol. 49, no. 3, pp. 626–634, March 2003.

[7] F. Baumgarte and C. Faller, "Binaural cue coding - Part II: Schemes and applications," *IEEE Trans. Speech Audio Processing*, vol. 11, no. 6, pp. 520–531, November 2003.

[8] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, revised ed. MIT Press, Cambridge, Massachusetts, 1997.

[9] B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hear Res.*, vol. 47, pp. 103–138, 1990.

[10] W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," *Journal of the Acoustical Society of America*, vol. 97, no. 6, pp. 3907–3908, May 1995.