USING A SINES + WAVELETS MIXED DICTIONARY FOR IMPROVING MATCHING PURSUIT-BASED PARAMETRIC AUDIO CODING

P. Vera-Candeas, N. Ruiz-Reyes, M. Rosa-Zurera, J.C. Cuevas-Martinez, J.M. Garcia-Rubia

Electronics and Telecommunication Engineering Department, Polytechnic School, University of Jaén
C/ Alfonso X el Sabio, 28, 23700 - Linares, Jaén, Spain
phone: +(34) 953648554, fax: +(34) 953648508, email: {pvera, nicolas,jccuevas}@ujaen.es

ABSTRACT

[illegible] MPEG-2/4-AAC [illegible] 16 [illegible]

1. INTRODUCTION

[illegible] [1, 2, 3, 4]. [illegible]

[illegible] STN [illegible] [3, 4, 5] [illegible]

[illegible] 1) [illegible] 2) [illegible] [1], [illegible]

[illegible] 1) [illegible] 2) [illegible]

2. SPARSE APPROXIMATIONS: MATCHING PURSUIT

[illegible] Mallat and Zhang [6].

Let [illegible] $x[n]$ [illegible] $g_i[n]$ [illegible]. Let $\mathbf{H}$ be a Hilbert space. We [illegible] $D = \{g_i[n]\ ; i = 0, 1, ..., L\}$ [illegible] $\mathbf{H}$, [illegible] $||g_i[n]|| = 1$.

[illegible] $x[n]$ [illegible] $g_i[n]$ [illegible] D, [illegible] l^2 [illegible] ([illegible] $g_i[n]$ [illegible] $x[n]$ [illegible]. At the m-th [illegible]

$$r^m[n] = r^{m+1}[n] + \alpha_{i(m)} \cdot g_{i(m)}[n] \quad m \geq 1 \tag{1}$$

[illegible] $\alpha_{i(m)}$ [illegible] $g_{i(m)}[n]$ [illegible] m-th [illegible] $r^1[n]$ [illegible] to $x[n]$.

[illegible] $r^m[n]$ [illegible] $g_i[n] \in D$, [illegible] m-th [illegible]

$$\alpha_i^m = \frac{\langle r^m[n], g_i[n]\rangle}{\langle g_i[n], g_i[n]\rangle} = \frac{\langle r^m[n], g_i[n]\rangle}{\|g_i[n]\|^2} = \langle r^m[n], g_i[n]\rangle \tag{2}$$

[illegible] $g_{i(m)}[n]$ [illegible] m-th [illegible]

$$g_{i(m)}[n] = \arg\min_{g_i \in D} \|r^{m+1}[n]\|^2 \tag{3}$$

From (1) and (2), [illegible] (3) [illegible]

$$g_{i(m)}[n] = \arg\max_{g_i \in D} |\alpha_i^m| \tag{4}$$

As [illegible] m-th [illegible] $r^m[n]$. [illegible] $\langle r^m[n], g_i[n]\rangle \forall g_i[n]$ [illegible] [6]:

$$\langle r^{m+1}[n], g_i[n]\rangle = \langle r^m[n], g_i[n]\rangle - \alpha_{i(m)} \cdot \langle g_{i(m)}[n], g_i[n]\rangle \tag{5}$$

[illegible] $\langle g_{i(m)}[n], g_i[n]\rangle$ [illegible] ($\langle x[n], g_i[n]\rangle$).

3. MATCHING PURSUIT WITH A MIXED DICTIONARY OF SINES + WAVELETS

[illegible] D [illegible] D_e [illegible] D_w. Let's $e_i[n]$ and $w_i[n]$ [illegible]

3.1 Implementation with sets of complex exponentials

[illegible]

$$e_i[n] = \frac{1}{\sqrt{N}} \cdot e^{j\frac{2\pi}{2L}n};\ i = 0,\dots,L-1;\ n = 0,\dots,N-1 \quad (6)$$

[illegible] $\frac{1}{\sqrt{N}}$ [illegible] N [illegible] L [illegible]

[illegible] (7):

$$\begin{aligned} r^{m+1}[n] &= r^m[n] - \alpha_{i(m)} e_{i(m)}[n] - \alpha^*_{i(m)} e^*_{i(m)}[n] \\ &= r^m[n] - 2Re\{\alpha_{i(m)} e_{i(m)}[n]\} \end{aligned} \quad (7)$$

[illegible] $r^m[n]$ [illegible] $e_i[n] \in D_e$, [illegible] $\{\alpha_i^m,\ i = 0,1,\dots,L-1\}$ [illegible]

$$\alpha_i^m = \frac{\langle r^m[n], e_i[n]\rangle - \langle r^m[n], e_i[n]\rangle^* \cdot \langle e_i^*[n], e_i[n]\rangle}{1 - |\langle e_i^*[n], e_i[n]\rangle|^2} \quad (8)$$

[illegible] α_i^m, [illegible] $e_{i(m)}$ [illegible] m- [illegible] (4).

[illegible] $e_i[n] \in D_e$ [illegible]

$$\begin{aligned} \langle r^{m+1}[n], g_i[n]\rangle &= \langle r^m[n], g_i[n]\rangle - \\ &- \alpha_{i(m)} \langle g_{i(m)}[n], g_i[n]\rangle - \\ &- \alpha^*_{i(m)} \langle g^*_{i(m)}[n], g_i[n]\rangle \end{aligned} \quad (9)$$

[illegible] $e_i[n] \in D_e$ [illegible] FFT [illegible] $x[n]$ [illegible] $e_i[n] \in D_e$ [illegible]

$$\langle x[n], e_i[n]\rangle = \frac{1}{\sqrt{N}} \cdot \sum_{n=0}^{2L-1} x[n] \cdot e^{-j\frac{2\pi}{2L}n} = \frac{1}{\sqrt{N}} \cdot X[i] \quad (10)$$

[illegible] $X[i]$ [illegible] $2L$- [illegible] DFT [illegible] $x[n]$, [illegible] $L > N$ [illegible] $e_i[n] \in D_e$ [illegible]

$$\begin{aligned} \langle e_{i(m)}[n], e_i[n]\rangle &= \frac{1}{N} \cdot \sum_{n=0}^{2L-1} e^{-j\frac{2\pi(i-i(m))}{2L}n} \\ &= \frac{1}{N} \cdot U[((i - i(m)))_{2L}] \end{aligned} \quad (11)$$

$$\begin{aligned} \langle e^*_{i(m)}[n], e_i[n]\rangle &= \frac{1}{N} \cdot \sum_{n=0}^{2L-1} e^{-j\frac{2\pi(i+i(m))}{2L}n} \\ &= \frac{1}{N} \cdot U[((i + i(m)))_{2L}] \end{aligned} \quad (12)$$

[illegible] $U[i]$ [illegible] $2L$- [illegible] DFT [illegible] $u[n]$. [illegible] 1) [illegible] $2L$- [illegible] FFT; 2) [illegible] $2L$- [illegible]

3.2 Implementation with sets of wavelet functions

[illegible] (16)). [illegible] $D = D_w$ [illegible] P- [illegible] (WP) [illegible] D_w [illegible] P- [illegible] WP [illegible] $\{s, p, k\}$, [illegible] s [illegible] p [illegible] k, [illegible] m- [illegible]

$$\alpha^m_{\{s,p,k\}} = \langle r^m[n], w_{\{s,p,k\}}[n]\rangle \quad (13)$$

$$w_{\{s,p,k\}}[n] = w_{\{s,p\}}[n - 2^p k] \quad (14)$$

[illegible] $\alpha^m_{\{s,p,k\}}$ [illegible] (13), [illegible] $w_{\{s,p,k\}(m)}[n]$ [illegible] m- [illegible]

$$w_{\{s,p,k\}(m)}[n] = \arg \max_{w_{\{spk\}} \in D_w} |\alpha^m_{\{s,p,k\}}| \quad (15)$$

[illegible] $\langle x[n], w_{\{s,p,k\}}[n]\rangle$ [illegible] $\langle w_{\{s_1,p_1,k_1\}}[n], w_{\{s_2,p_2,k_2\}}[n]\rangle$. [illegible] WP [illegible] $x[n]$. [illegible] 7]:

$$\langle w_{\{s_1,p_1,k_1\}}, w_{\{s_2,p_2,k_2\}}\rangle = \begin{cases} \delta[k_2 - k_1] & s_1 = s_2,\ p_1 = p_2 \\ 0 & s_2 \neq \lfloor \frac{s_1}{2^{p_1 - p_2}} \rfloor \\ w_{\{s,p\}}[k_2 - 2^p k_1] & s_2 = \lfloor \frac{s_1}{2^{p_1 - p_2}} \rfloor \end{cases} \quad (16)$$

[illegible] $p = p_1 - p_2$, $s = ((s_1))_{2^p}$ and $p_1 \geq p_2$ [illegible] (16), [illegible] WP [illegible]

3.3 Implementation with the mixed dictionary

[illegible] $D = D_e + D_w$), [illegible] $\{\alpha_i^m, \alpha^m_{\{s,p,k\}}\}$ [illegible] $\{e_i[n], w_{\{s,p,k\}}[n]\}$ [illegible] (8) [illegible] (13), [illegible]

[illegible]

3.3.1 *Correlation between two complex exponentials.*

[illegible] FFT. [illegible] $e_i[n] \in D_e$ [illegible] (11) [illegible] (12) [illegible] 3.1. [illegible] $e_i[n]$, [illegible] (9) [illegible]

3.3.2 *Correlation between a complex exponential and a wavelet function when the complex exponential is chosen.*

[illegible] (7). [illegible] $e_{i(m)}[n]$ [illegible] $w_{\{s,p,k\}}[n] \in D_w$ [illegible] (9). [illegible] DFT [illegible]

$$\langle e_{i(m)}[n], w_{\{s,p,k\}}[n]\rangle = \sum_{n=0}^{N-1} \frac{1}{\sqrt{N}} e^{j\frac{2\pi\ (m)}{2L}n} w_{\{s,p,k\}}[n] = \frac{1}{\sqrt{N}} W^*_{\{s,p,k\}}[i(m)] \qquad (17)$$

[illegible] $W_{\{s,p,k\}}[i(m)]$ [illegible] 2 L-[illegible] DFT [illegible] $w_{\{s,p,k\}}[n]$ [illegible] $\frac{i(m)}{2L}$. [illegible] 2 L-[illegible] DFT [illegible] $w_{\{s,p,k\}}[n]$ [illegible] $w_{\{s,p,k\}}[n] = w_{\{s,p\}}[n-2^pk]$, [illegible] 2 L-[illegible] DFT [illegible] $w_{\{s,p\}}[n]$. [illegible] 2 L-[illegible] DFT [illegible] DFT.

3.3.3 *Correlation between two wavelet functions.*

[illegible] D_w [illegible] $w_{\{s,p,k\}}[n] \in D_w$ [illegible] (16). [illegible] (16), [illegible] WP [illegible]

3.3.4 *Correlation between a complex exponential and a wavelet function when the wavelet function is chosen.*

[illegible] $w_{\{s,p,k\}(m)}[n]$ [illegible] $e_i[n] \in D_e$:

$$\langle w_{\{s,p,k\}(m)}[n], e_i[n]\rangle = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} w_{\{s,p,k\}(m)}[n] e^{-j\frac{2\pi}{2L}n} = \frac{1}{\sqrt{N}} W_{\{s,p,k\}(m)}[i] \qquad (18)$$

[illegible] $W_{\{s,p,k\}(m)}[i]$ [illegible] 2 L-[illegible] DFT [illegible] $w_{\{s,p,k\}}[n]$ [illegible] $\frac{i}{2L}$. [illegible] 2 L-[illegible] DFT [illegible] $w_{\{s,p,k\}}[n]$ [illegible] 3.3.2, [illegible] $w_{\{s,p,k\}}[n] = w_{\{s,p\}}[n-2^pk]$.

[illegible] ($D = D_e + D_w$) [illegible] 1) [illegible] 2 L-[illegible] DFTs [illegible] $e_i[n] \in D_e$; 2) [illegible] WP [illegible] ($w_{\{s,p\}}[n]$); 3) [illegible] 2 L-[illegible] DFTs [illegible] $w_{\{s,p\}}[n] \in D_w$.

4. EXPERIMENTAL RESULTS

[illegible] (1) [illegible] (2) [illegible] (3) [illegible]

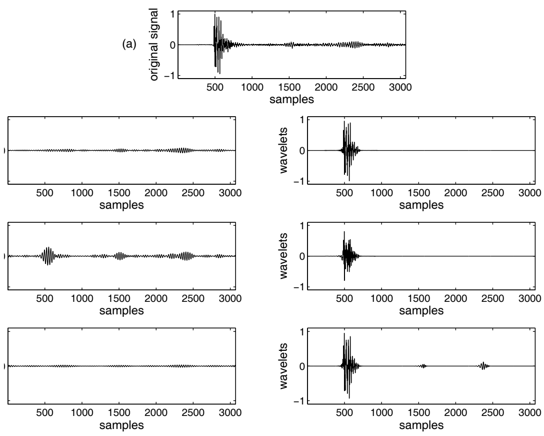

Fig. 1: *(a) Audio fragment containing a signal onset. (b) Sinusoids and wavelets extracted from the audio frame by the first approach. (c) The same components extracted by the second approach. (d) Idem by the third approach.*

[illegible] 2% [illegible]

[illegible] 2 [illegible] 1. [illegible]

[illegible] STN-[illegible]

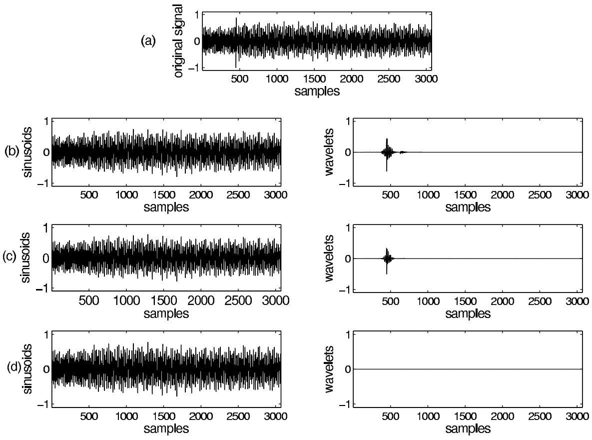

Fig. 2: *Audio fragment containing a micro-transient*

[illegible] MPEG [illegible] 50% [illegible] 23 - [illegible] $(N = 1024)$ and $L = 4096$. [illegible] OAB [illegible] O [illegible] A and B [illegible] (A or B) [illegible] 1.

Table 1: PREFERENCE FOR MIXED DICTIONARY-VS-CASCADED DICTIONARIES (%).

[illegible]	Pref. (%)
[illegible]	100
[illegible]	100
[illegible]	52
[illegible]	46
[illegible]	100
[illegible]	70
[illegible]	56
[illegible]	60
[illegible]	100

[illegible] 1 [illegible]

[illegible] [4]. Fig. 3 [illegible] MPEG-2/4 - AAC at 16 Kbits/s [illegible] [4]. [illegible] MUSHRA [8] [illegible]

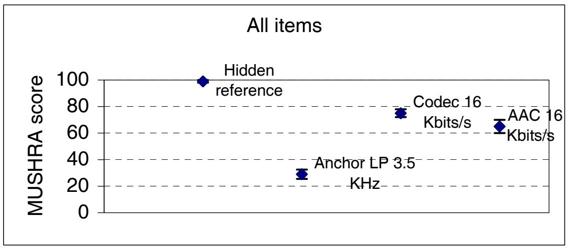

Fig. 3: *MUSHRA listening test results showing mean grading and 95% confidence interval.*

It [illegible] MPEG-2/4 - AAC at 16 Kbits/s

5. CONCLUSIONS

[illegible]

[illegible] MPEG-2/4 - AAC at 16 Kbits/s

REFERENCES

[1] S. L[illegible] and J. S[illegible], "A S[illegible] [illegible]," Preprint 4781, *105th AES Convention*, San Francisco (USA), 1998.

[2] H. P[illegible] and N. M[illegible], "HILN - [illegible] MPEG-4 [illegible]," *Proc. ISCAS*, [illegible], vol. 3, pp. 201-204, 2000.

[3] E. S[illegible], "[illegible]," *114th AES Convention: Workshop on recent developments in MPEG-4 audio*, March 2003.

[4] P. V[illegible], N. R[illegible], M. R[illegible], J. C[illegible] and P. J. R[illegible], "[illegible]," Preprint 6176, *116th AES Convention*, Berlin (Germany), 2004.

[5] T. S. V[illegible] and T. H. Y. M[illegible], "A 6 [illegible] 85 [illegible]," *Proc. ICASSP*, [illegible], pp. 877-880, 2000.

[6] S. Mallat and Z. Zhang, "Matching Pursuit [illegible]," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3397-3415, Dec. 1993.

[7] P. V[illegible], N. R[illegible], M. R[illegible], D. M[illegible] and F. L[illegible], "[illegible]," *IEEE Signal Processing Letters*, vol. 11, no. 3, 2004.

[8] ITU-R, "[illegible] (MUSHRA)", *ITU-R Recommend.*, BS. 1534, 2001.