# ANALYSIS / SYNTHESIS OF ROLLING SOUNDS USING A SOURCE-FILTER APPROACH

*Jung Suk Lee, Philippe Depalle and Gary Scavone*

Music Technology
Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT)
Schulich School of Music of McGill University
Montreal, QC, Canada
jungsuk.lee@mail.mcgill.ca

## ABSTRACT

In this paper, the analysis and synthesis of a rolling ball sound is proposed. The approach is based on the assumption that the rolling sound is generated by a concatenation of micro-impacts between a ball and a surface, each having associated resonances. Contact timing information is first extracted from the rolling sound using an onset detection process. The resulting individual contact segments are subband filtered before being analyzed using linear predictive coding (LPC) and notch filter parameter estimation. The segments are then resynthesized and overlap-added to form a complete rolling sound. This approach is similar to that of [1], though the methods used for contact event detection and filter parameter estimation are completely different.

## 1. INTRODUCTION

The synthesis of rolling sounds has applications in virtual reality and the game industry, where high-quality parametric models of environmental sounds are important in creating a realistic and natural result. Methods for the synthesis of rolling sounds have been studied by several researchers. Van den Doel [2] proposed a source-filter approach to produce rolling sounds using colored noise as input to a resonant filter structure. In [3] and [4], a real-time physics-based parametric 'cartoonification' model of a rolling object was proposed, with the seemingly continuous interactions of a ball and surface approximated as a sequence of distinct ball-surface contact events. They developed a non-linear physical impact model and incorporated that with a modal synthesis technique. Lagrange et al. [1] assumed a similar ball-surface interaction but focused on a linear source-filter model for the analysis and resynthesis. They used a high-resolution method to estimate a fixed resonant filter characteristic and an iterative contact event estimation method based on fitting a parameterized impact window to the sound to determine the source signal.

In this paper, we assume the rolling sound is composed of a collection of micro-collisions between a rolling object and an uneven surface (as in [3] and [1]). We first perform a contact event estimation and segment the sound accordingly. We then estimate a separate filter characteristic for each segment. In this way, it is possible to account for the varying modal property with respect to the locations of the contacts along the trajectory of the object on the surface, which in turn allows a physically intuitive analysis. We first describe how to decompose the rolling sound signal into individual contact segments. Secondly, the analysis and synthesis of each segment are discussed. A filter bank splits each segment into subbands with different frequency bandwidths, which enables a better LPC estimation, especially for strong low-frequency resonant modes. We also perform a notch filter estimation to account for effects related to position-dependent excitation of the modes of a plate.

## 2. DETECTION OF CONTACT TIMINGS AND THE DEFINITION OF ONE CONTACT SOUND

Our approach is based on the underlying assumption that the rolling sound results from numerous micro-collisions of a rolling object over an uneven surface. Thus, we must find the contact timings so that analyses of individual contact events can be accomplished. To this end, high-pass filtering is first performed on the signal to help distinguish contacts by their high-frequency transients. Fig.1 shows a rolling sound signal, denoted as $y(n)$, its high-pass filtered version (cutoff frequency is 10kHz with sampling frequency 44.1kHz), $y_{hp}(n)$, and the spectrogram of $y(n)$.

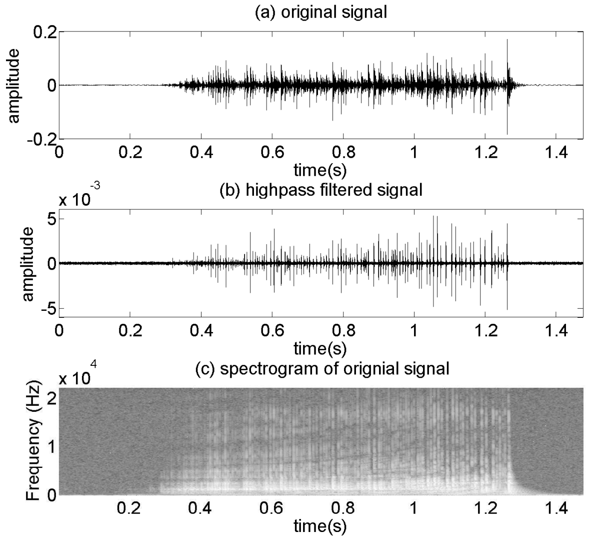

Figure 1: Top: Original rolling sound $y(n)$. Middle: High pass filtered rolling sound $y_{hp}(n)$. Bottom: Spectrogram of $y(n)$.

A linear-phase high-pass filter is used to obtain $y_{hp}(n)$ so that group delays can be easily aligned. In order to detect contact timings more accurately, an onset detection process is performed on

$y_{hp}(n)$. An envelope function $E(n)$ is defined as in Eq. (1) [5] and the *box function* $b(n)$ is defined as in Eq. (2) by replacing all values of $E(n)$ greater than some threshold with a positive value $\alpha$.

$$E(n) = \frac{1}{N}\sum_{m=-\frac{N}{2}}^{\frac{N}{2}-1}[y_{hp}(n+m)]^2 \quad (1)$$

$$b(n) = \begin{cases} \alpha & \text{if } \quad E(n) > \text{threshold} \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Fig. 2(a) shows $E(n)$ of $y(n)$ and Fig. 2(b) shows an enlarged portion of $E(n)$ and its associated $b(n)$. In order to detect the onset times from $b(n)$, $d(n)$, a time derivative of $b(n)$, is computed using a simple differencing operation (high-pass filtering) as given by Eq. (3). As shown in Fig. 2(c), $d(n)$ contains values of either $\alpha$ or -$\alpha$.

$$d(n) = b(n) - b(n-1) \quad (3)$$

$$o(n) = \begin{cases} d(n) & \text{if} \quad d(n) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Finally, the value for $o(n)$ given by Eq. (4) is obtained by rejecting the negative values in $d(n)$, as shown in Fig. 2(d), from which the contact timing index information function $i(k)$ is finally defined as

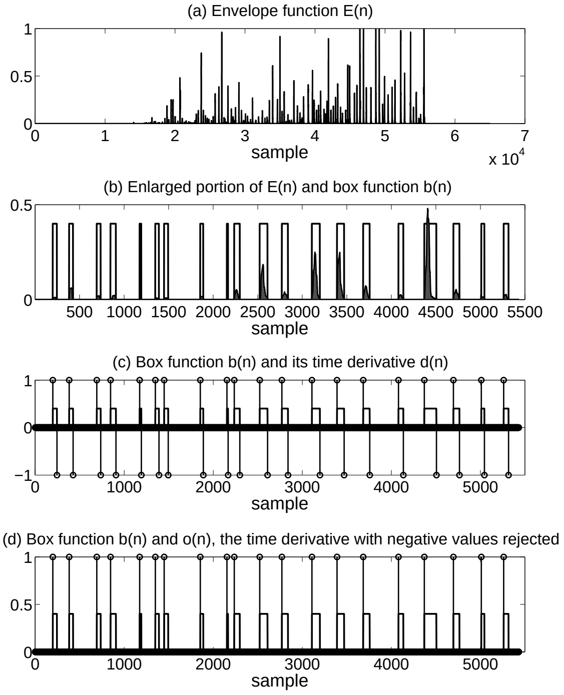

Figure 2: (a) Envelope function $E(n)$ of given rolling sound $y(n)$. (b) Enlarged portion of $E(n)$ in (a) and its associated box function $b(n)$ (eq. (2)). (c) Box function $b(n)$ from (b) and $d(n)$, a time derivative of $b(n)$ (vertical lines) (eq. (3)), here $\alpha = 1$. (d) Box function $b(n)$ from (b) and $o(n)$ (vertical lines).

$$i(k) \; : \; \text{sample indexes where } o(n) = \alpha \quad (5)$$
$$k = 1, 2, 3, 4, \cdots, N_\alpha$$

where $N_\alpha$ is the total number of $\alpha$ in $o(n)$. From our basic assumption of the rolling dynamics, we wish to feed one contact sound at a time into the analysis/synthesis system. We thus define 'one contact sound' as a segment of the original signal $y(n)$ whose length is the interval between two identified adjacent contact indices $i(k+1)-i(k)$. The $k$th contact sound $x_k(n)$ is defined as follows:

$$x_k(n) = y(n + i(k) - 1), \; n = 1, 2, ..., i(k+1) - i(k) \quad (6)$$
$$k = 1, 2, \cdots, N_\alpha$$

## 3. ANALYSIS AND SYNTHESIS SYSTEM

The analysis and synthesis scheme proposed here is devised to identify the excited modes of a single contact sound $x_k(n)$. An input segment $x_k(n)$ is first decomposed into subband signals by a tree structure filter bank. This not only improves the LPC analysis by limiting the frequency range over which resonances are estimated but it also allows for different LPC parameters in each subband, perhaps informed by perceptual characteristics. We also perform a notch detection operation for each segment to account for the time-varying, position-dependent suppression/attenuation of resonant modes as an object rolls over a surface.

### 3.1. Tree structure filter bank

A tree structure filter bank [6] is used to separate each contact segment into different frequency bands having unequal bandwidths for both the analysis and synthesis operations. The tree structure filter bank is constructed with two basic filters – one lowpass filter and another high pass filter – and two-channel quadrature mirror filters (QMF) (Fig. 3) are used to achieve a perfect reconstruction (PR) condition for the filter bank. Four filters of the QMF bank at the analysis bank (AB), $H_0(z)$, $H_1(z)$, and the synthesis bank (SB), $G_0(z)$, $G_1(z)$, are related as below, from which the alias-free property and the power symmetric condition are met [6].

$$H_1(z) = H_0(-z), \; G_0(z) = H_0(z), \; G_1(z) = -H_1(z) \quad (7)$$

A 4-band structure was empirically chosen, with cut-off frequencies of $\frac{1}{8}\pi$, $\frac{1}{4}\pi$, $\frac{1}{2}\pi$ on a normalized frequency axis (Fig. 5). This filter bank can also be represented as a typical 4-band filter bank (Fig. 4) using the noble identity [6]. The first band corresponding to the lowest frequency subband $V_1(z)$ at the AB is equivalent to $H_0(z)H_0(z^2)H_0(z^4)$, and the rest of the bands are denoted such that $V_2(z) = H_0(z)H_0(z^2)H_1(z^4)$, $V_3(z) = H_0(z)H_1(z^2)$, $V_4(z) = H_1(z)$. In the same way, the filters at the SB $W_l(z)$, $l = 1, 2, 3, 4$ are defined. In addition, the filter $H_0(z)$ is designed with a linear phase characteristic [7] so that the whole tree structure filter bank is piecewise linear phase. Therefore we are able to easily compensate for group delays introduced by the filter bank with simple time-domain shifting.

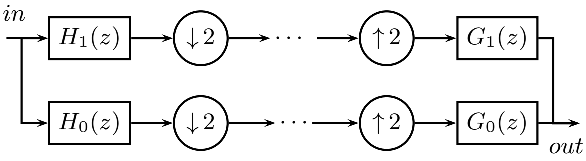

Figure 3: Two channel QMF bank.

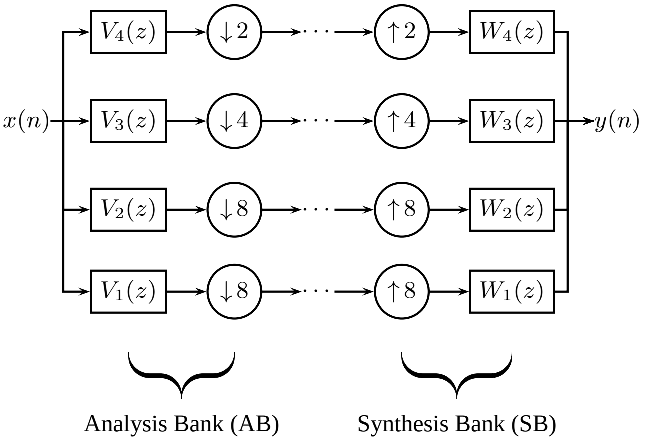

Figure 4: Non uniform 4-band filter.

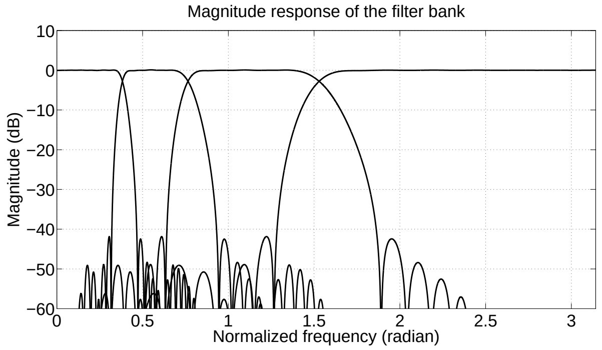

Figure 5: Magnitude response of the filter bank.

### 3.2. Analysis/synthesis of one contact sound

When an object collides with a surface, the surface is set into motion by the excitation force exerted by the object. The vibrational motion of the surface is characterized by its modal properties, which in turn are determined by its geometry and physical characteristics. Due to the finite dimension of the surface, modes are selectively excited and attenuated / suppressed depending on the location of the excitation. In a frequency magnitude response, excited modes appear as peaks and suppressed modes appear as time-varying notch patterns that move in a self-consistent way over regions of the spectrum where energy was previously found. For example, Fig.6 illustrates a simulated modal pattern for contacts along the length of a simply supported rectangular plate and a similar upwardly varying notch pattern in the spectrogram of $y(n)$. For each contact segment, we thus model both the time-varying spectral peaks (using LPC) and the notches. In general, it is known that excited modes, which represent resonances, are essential for the perception of the sounds. However, in the case of rolling sound on a finite length rigid surface, existing notches in the spectrum are also important as their notch-frequencies vary with time. In addition, by estimating notches, we can reduce the LPC order (which would otherwise be unnecessarily high to describe zeros[8]).

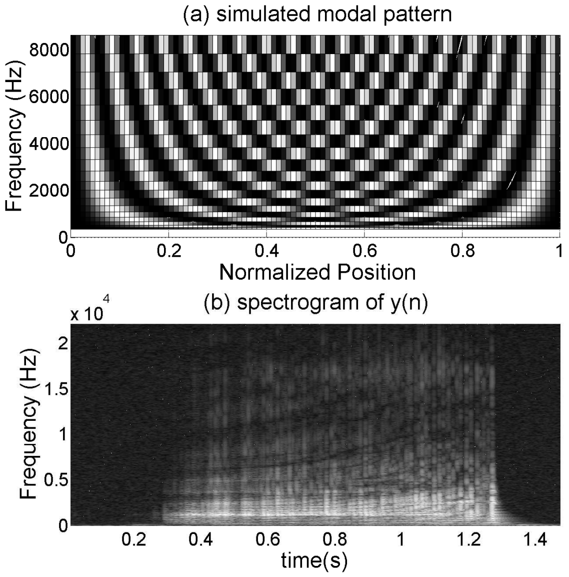

Figure 6: (a) Simulated modal pattern of simply supported rectangular MDF plate (Width: 0.95(m), height: 0.25(m), thickness: 0.02(m)) excited by a rolling object traveling from one end to the other end of the longer side, while centered on the other axis. (b) Spectrogram of $y(n)$. Upwardly varying notches can be seen.

#### *3.2.1. Estimation of zeros using notch filtering*

The $k$th input signal $x_k(n)$ is split into 4 subbands and downsampled at AB. Subband signals $x_k^{(l)}(n)$ are defined as below:

$$x_k^{(l)}(n) = \textsc{Downsample}(v_l(n) * x_k(n)), \tag{8}$$

where $l$ is the order of the subband and $v_l(n)$ is the impulse response of $l$th subband filter at the AB of the filter bank. DOWNSAMPLE denotes a downsampling operation.

Because the attenuated modes, as well as the excited modes, are perceptually important in characterizing by the location of the rolling object, we focus on the estimation of the suppressed modes, represented as notches in the spectrum, as well as the excited modes.

In order to estimate notches of $x_k^{(l)}(n)$, we considered building a notch filter where frequencies and bandwidths of notches are modeled according to the valleys in the frequency response of $x_k(n)$. To this end, $|X_k^{(l)}(e^{j\omega})|$ (Fig. 7(a)), the magnitude of the Fourier Transform of $x_k^{(l)}(n)$, is flipped to $1/|X_k^{(l)}(e^{j\omega})|$ (Fig. 7(b)) and peak frequencies of it ($\omega_{k,m}^{(l)}$, for $m$=1,2,. . .,$M$, where $M$ is the number of detected peaks) are detected using the Matlab function `findpeaks`. Then by using quadratic polynomial curve fitting, lobes representing peaks are modeled to estimate 3dB-bandwidths $BW_{k,m}^{(l)}$ [9]. Once $\omega_{k,m}^{(l)}$ and $BW_{k,m}^{(l)}$ are estimated (normalized radian frequencies), we can form a zero as $z_{k,m}^{(l)} = e^{(-BW_{k,m}^{(l)}/2)} e^{-j\omega_{k,m}^{(l)}}$ representing a valley in $|X_k^{(l)}(e^{j\omega})|$ [9]. Then biquad sections representing a suppressed mode are derived as

$$\begin{aligned} B_{k,m}^{(l)}(z) &= (1 - z_{k,m}^{(l)} z^{-1})(1 - \overline{z}_{k,m}^{(l)} z^{-1}) \qquad (9) \\ A_{k,m}^{(l)}(z) &= (1 - \rho z_{k,m}^{(l)} z^{-1})(1 - \rho \overline{z}_{k,m}^{(l)} z^{-1}), \qquad (10) \end{aligned}$$

where $\rho = 0.95$ and $\overline{z}_{k,m}^{(l)}$ denotes the complex conjugate of $z_{k,m}^{(l)}$. $A_{k,m}^{(l)}$, a biquad section which is the denominator of the notch filter, is necessary to isolate each notch properly [10]. The notch filter $N_k^{(l)}(z)$ is given as follows:

$$N_k^{(l)}(z) = \prod_{m=1}^{M} \frac{B_{k,m}^{(l)}(z)}{A_{k,m}^{(l)}(z)}. \tag{11}$$

As shown in Fig. 7(c), the constructed notch filter has notches whose frequencies and bandwidths are the same as those of peaks in $1/|X_k^{(l)}(e^{j\omega})|$ (Fig. 7(b)). $X_k^{(l)}(z)$ is then filtered with $1/N_k^{(l)}(z)$ as below to obtain $Q_k^{(l)}(z)$:

$$Q_k^{(l)}(z) = \frac{X_k^{(l)}(z)}{N_k^{(l)}(z)}. \tag{12}$$

In $Q_k^{(l)}(e^{j\omega})$, notches are removed since $1/N_k^{(l)}(z)$ is an inverse filter of the notch filter, thus enabling LPC estimation with lower orders (Fig. 7(d)).

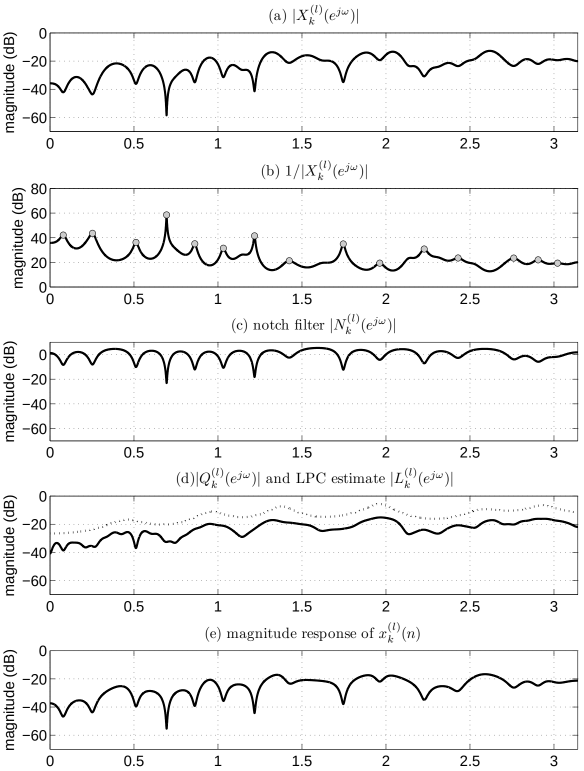

Figure 7: (a) Magnitude of $X_k^{(l)}(e^{j\omega})$. (b) Magnitude of $1/X_k^{(l)}(e^{j\omega})$. Circle marks denote detected peaks. (c) Magnitude response of the notch filter $N_k^{(l)}(z)$. (d) Magnitudes of $Q_k^{(l)}(z)$ (solid) and its LPC estimate $L_k^{(l)}(z)$ (dashed). (e) Magnitude of $\hat{X}_k^{(l)}(z)$. In all figures, $x$-axes denote normalized radian frequencies.

### 3.2.2. Estimation of poles using LPC

In order to estimate poles from $Q_k^{(l)}(z)$, a $p_l$th-order LPC estimate $L_k^{(l)}(z)$ is derived as below:

$$L_k^{(l)}(z) = \frac{G_k^{(l)}}{1 - \sum_{m=1}^{p_l} a_m^{(l,k)} z^{-m}}, \tag{13}$$

where $L_k^{(l)}(z)$ is the transfer function, $a_m^{(l,k)}$ are LPC coefficients and $G_k^{(l)}$ is a gain of the LPC estimate. $a_m^{(l,k)}$ are estimated in such a way that the linear prediction error $e_k^{(l)}(n)$ as defined below is minimized [11].

$$e_k^{(l)}(n) = q_k^{(l)}(n) - \sum_{m=1}^{p_l} a_m^{(l,k)} q_k^{(l)}(n-m), \tag{14}$$

where $q_k^{(l)}(n)$ is the impulse response of the $Q_k^{(l)}(z)$. $\hat{X}_k^{(l)}(z)$, the synthesis result of $X_k^l(z)$, is finally derived as follows:

$$\hat{X}_k^{(l)}(z) = W_1(z)(\text{UPSAMPLE}(L_k^{(l)}(z)N_k^{(l)}(z))) \tag{15}$$

where UPSAMPLE denotes an upsampling operation. LPC order $p_l$ varies along subbands.

In Fig. 8, the LPC estimates of the subband signals and the fullband signal used for the example of Fig. 7 are shown. All magnitude responses shown in Fig. 8 are without zero estimates applied. In the example of Fig. 7 and Fig. 8, the length of the contact sound $x_k(n)$ is 490 sample and the sampling rate is 44.1kHz. LPC orders are set to $p_1 = 25$, $p_2 = 10$ and $p_3 = p_4 = 5$ for the subband signals and 45 for the fullband signal (no filter bank applied) so that the total orders of both cases are the same. It is clear that as a higher order is used for the low frequency region, significant spectral peaks are more effectively handled with limited orders.

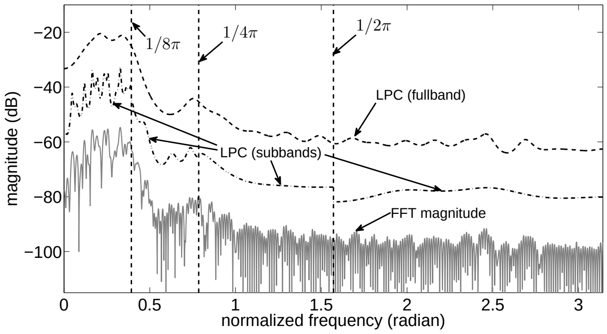

Figure 8: Magnitudes of $X_k(z)$ and its syntheses. Gray line is the magnitude plot of $X_k(e^{j\omega})$ and black dotted line is the full band LPC estimate with order 45. Black dash-dotted lines are the magnitude responses of the LPC estimates of the subbands signal from the lowest subband to the highest subband, respectively (Zero estimates are not considered). LPC orders are 25, 10, 5, 5, from the lowest to the highset, respectively.

Since all the subband filters employed in the SB are linear phase, their group delays $\tau_l^{(k)}$ are frequency independent and only simple time-domain shifts arise as phase distortion. This well behaved group delay property of the analysis/synthesis system is clearly evident in Fig. 9. Therefore, the phase distortion of $\hat{x}_k^{(l)}(n)$, the impulse response of $\hat{X}_k^{(l)}(z)$, can be easily adjusted by shifting $\hat{x}_k^{(l)}(n)$ by $\tau_l^{(k)}$ which is estimated from the filter orders of $W_l(z)$. To complete the synthesis of $x_k(n)$, $\hat{x}_k^{(l)}(n)$ are shifted back by $\tau_l$ and added together as illustrated below to form $s_k$, the final synthesis output corresponding to $x_k(n)$.

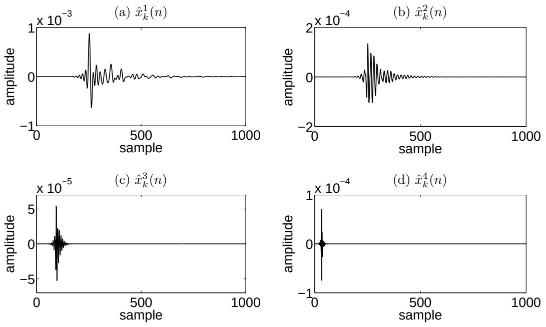

Figure 9: Synthesized subband outputs ($\hat{x}_k^l(n)$). (a) $\hat{x}_k^{(1)}(n)$, (b) $\hat{x}_k^{(2)}(n)$, (c) $\hat{x}_k^{(3)}(n)$, (d) $\hat{x}_k^{(4)}(n)$.

$$\hat{x}_k^{(l)}(n) = \hat{x}_k^{(l)}(n + \tau_l^{(k)} - 1) \qquad (16)$$

$$s_k(n) = \sum_{l=1}^{num} \hat{x}_k^{(l)}(n) \qquad (17)$$

$num$ : total number of subbands

As the transfer function of the synthesis result $\hat{X}_k^{(l)}(z)$ includes infinite impulse response (IIR) components, the impulse response $\hat{x}_k^{(l)}(n)$ must be truncated so as to make the length of $s_k(n)$ finite. The $s_k(n)$ segments are thus cascaded using the overlap and add method in such a way that the location of $s_k(1)$ is matched to $i(k)$. Thus the tail of $s_k(n)$ is overlapped with a part of $s_{k+1}(n)$ and then added together.

Sound examples are available at:
http://www.music.mcgill.ca/∼lee/DAFx10

## 4. CONCLUSION

An analysis and synthesis approach for rolling sounds is proposed in this paper. The process is based on the assumption that the overall sound can be linearly decomposed into many micro-contacts between the object and the surface on which it rolls. Therefore, a process similar to onset detection is carried out to extract the contact timing information and segment the sound into individual contact events. Each segment is fed into an analysis/synthesis system to estimate time-varying filters. The analysis/synthesis process consists of a tree structure filter bank, LPC processors and notch filters. Thanks to the tree structure filter bank, LPC orders can be flexibly assigned to subbands, allowing us to focus more on significant spectral features while analyzing and synthesizing with LPC processors and notch filters. Finally, the resynthesized contact events are appropriately cascaded using the overlap and add method to produce the final rolling sound.

## 5. REFERENCES

[1] M. Lagrange, G. Scavone, and P. Depalle, "Analysis/synthesis of sounds generated by sustained contact between rigid objects," *IEEE Trans. Audio, Speech and Language Signal Processing*, vol. 18, no. 3, pp. 509–518, 2010.

[2] K. Van den Doel, P. G. Kry, and D. K. Pai, "Foleyautomatic:physically based sound effects for interactive simulation and animation," in *Proc. Int. Conf. Comput. Graphics and Interactive Techniques (SIGGRAPH)*, 2001.

[3] M. Rath, "An expressive real-time sound model of rolling," in *Proceedings of the 6th International Conference on Digital Audio Effects (DAFx-03)*, London, UK, Sept., 2003.

[4] D. Rocchesso and F. Fontana, *The Sounding Object*, Edizioni di Mondo Estremo, 2003.

[5] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A tutorial on onset detection in music signals," *IEEE Trans. on speech and audio processing*, vol. 13, no. 5, pp. 1035–1047, Sept. 2005.

[6] P. P. Vaidyanathan, *Multirate systems and filter banks*, Prentice Hall, 1993.

[7] T. Q. Nguyen and P. P. Vaidyanathan, "Two-channel perfect-reconstruction FIR QMF structures which yield linear-phase analysis and synthesis filters," *IEEE Trans. Acoust., Speech and Signal Processing*, vol. 37, no. 5, pp. 676–690, May 1989.

[8] S. M. Kay and S. L. Marple Jr, "Spectrum analysis : a modern perspective," *Proceedings of the IEEE*, vol. 69, no. 11, pp. 1380–1419, 1981.

[9] J. O. Smith, *Physical Audio Signal Processing*, W3K Publishing, `http://books.w3k.org`, 2007.

[10] D. B. Rao and S. Y. Kung, "Adaptive notch filtering for the retrieval of sinusoids in noise," *IEEE Trans. Acoust., Speech and Signal Processing*, vol. 32, no. 4, pp. 791–802, 1984.

[11] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.