POZNAN UNIVERSITY OF TECHNOLOGY ACADEMIC JOURNALS
No 54 Electrical Engineering 2007

Maciej BARTKOWIAK*
Tomasz ŻERNICKI*

IMPROVED PARTIAL TRACKING TECHNIQUE FOR SINUSOIDAL MODELING OF SPEECH AND AUDIO

The paper presents a simple, yet very effective tracking technique for time-varying sinusoidal partials within sinusoidal and hybrid models of speech and audio. The issues of several existing tracking techniques are discussed in the paper. We extend the classic tracking algorithm of McAulay and Quatieri by the inclusion of frequency and amplitude skew factors estimated in the process of chirp analysis. A simple measure of frequency and amplitude matching is proposed to support reliable pairing of detected quasi-sinusoidal peaks across successive frames. Moreover, the paper proposes an evaluation methodology for quality assessment of tracking algorithms.

Keywords: sinusoidal model, partial tracking, chirps, estimation, tracking quality

1. INTRODUCTION

Sinusoidal model [1,2], as well as several derived hybrid models [3], are well-established signal analysis transformation and synthesis tools in the field of speech and audio processing. In general, the deterministic part of the signal is modeled as an additive mixture of quasi-sinusoidal components whose parameters are time-varying (1).

$$x_{\mathrm{det}}(t) = \sum_{k=1}^{K} A_k(t) \cos\left(\varphi_k + 2\pi \int_0^t f_k(\tau)\, d\tau \right),. \tag{1}$$

The components are called sinusoidal partials since they locally resemble pure sinusoids. Nevertheless, the underlying amplitude and frequency modulations resulting from the intensity and pitch envelopes of the sound cause the spectral energy of each partial to be spread over certain frequency range. The partial parameters (amplitudes $A_k(t)$ and frequencies $f_k(t)$) are estimated from the original signal by the use of time-frequency analysis and tracked in consecutive time instants thus forming the frequency and amplitude trajectories (fig. 1).

* Poznan University of Technology.

In most implementations the signal is divided into overlapping frames and the (f_k, A_k) parameters are estimated once per frame with the assumption that they represent the mean value or the instantaneous value sampled at the center of the frame. In the process of signal reconstruction the intermediate frequency and amplitude values are approximated by simple interpolating functions (usually low order piecewise polynomials).

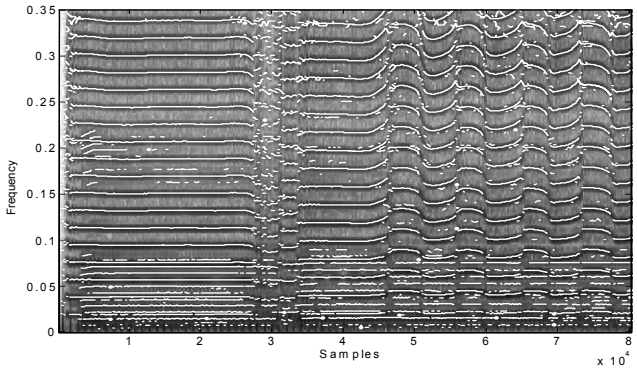

Fig. 1. Typical trajectories of sinusoidal partials shown on the background of signal spectrogram.

The key components that contribute to the accuracy of the sinusoidal model are partial estimation and tracking. Several robust and reliable techniques have been developed for the estimation of partial parameters [4], however perfect tracking of their evolution in time and frequency still appears to be a significant problem, especially in the context of modeling of polyphonic music [5]. In this paper we propose a relatively simple extension of the existing tracking algorithm originally proposed by McAulay and Quatieri [1]. The advantage of our approach is that it allows for tracking of partials with deep amplitude and frequency modulations. Such modulations are especially observed in the high order harmonics whose absolute frequency deviations in consecutive analysis frames are many times deeper than corresponding deviations of their fundamental frequency (cf Fig. 1). As opposed to other tracking techniques proposed hitherto, our algorithm does not make any heuristic assumptions about the consistency of underlying modulations. Furthermore, it operates online (frame by frame), which makes it possible to implement a low delay audio processing or coding system without any excessive delay resulting from long-term modeling of trajectories.

2. SINUSOIDAL MODELING ISSUES

Detection and estimation of partials, and partial tracking are usually two consecutive and mutually independent tasks in the analysis stage of the sinusoidal

model. First, classification of local peaks in the short-term spectra is performed, which involves making decisions for each peak, whether it represents a sinusoidal component or not (fig.2).

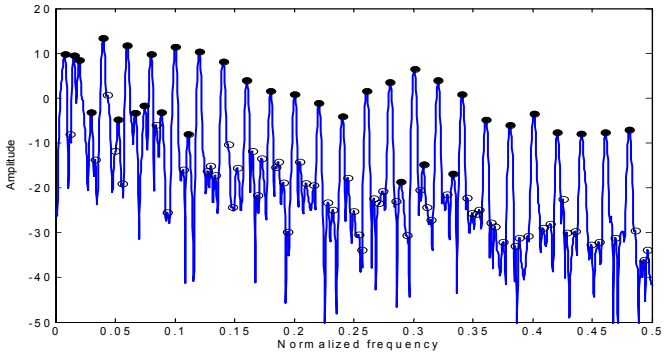

Fig. 2. Classification of spectral peaks in the short-term spectra. Filled circles denote peaks classified as corresponding to sinusoidal components

The list of detected sinusoidal peaks together with their estimated frequencies and amplitudes is passed to the tracking algorithm whose aim is to match corresponding peaks across consecutive analysis frames. While this matching is relatively easy in the case of single sounds of stable pitch (plucked strings, wind instruments), it may be extremely difficult in the case of heavily modulated sounds with unstable harmonics (bowed strings, singing voice), especially in the case of polyphonic music, where harmonics of different sounds often overlap and cross.

However correctly motivated by fundamental differences between detection and pattern analysis, separation of peak detection and tracking brings in fact many significant problems. The basic one is that once a wrong decision is made at the detection stage, the partial tracking algorithm must cope with incomplete or erroneous data.

In general, tracking always assumes some continuity in frequencies and amplitudes of corresponding partials due to significant frame overlap. Partials linked across consecutive frames define sinusoidal trajectories. Since reconstruction of the signal from model involves continuous parameter interpolation along each trajectory, proper connection of corresponding peaks is of crucial importance.

3. EXISTING PARTIAL TRACKING TECHNIQUES

The classical sinusoidal model developed by McAulay and Quatieri for analysis and resynthesis of monophonic speech is based on several simplifying assumptions. First of all, all local maxima of the STFT magnitude (including those

related to noise as well as spurious peaks resulting from analysis window) are detected and subject to tracking. Since all spectral components are intentionally represented as sinusoidal partials, the model generates an excessive amount of partial trajectories, which are often very short and chaotic. Nevertheless, partial matching strategy relies on the observation that speech contains at most one quasi-periodic component, hence all corresponding partial trajectories are approximately parallel in the time-frequency plane and do not cross. The tracking algorithm operates frame-wise. For each spectrum peak detected in the current frame it seeks a corresponding peak in the next frame, whose absolute frequency distance is the smallest and less then some allowed threshold Δ_f (2), yet it does not allow for crossing nor multiple connections between peaks.

$$f_k(t+\mathrm{T}) = \min_{f_k(t+T)} \left| f_k(t+\mathrm{T}) - f_k(t) \right| \Leftrightarrow \left| f_k(t+\mathrm{T}) - f_k(t) \right| < \Delta_f \tag{2}$$

Conflicts are resolved in favor of the closest frequency match, and the remaining peaks are forced to seek for a different match that still satisfies the absolute difference condition (Fig. 3). In the case of no matching peak in the current frame or in the subsequent frame, a corresponding track birth or death is marked for that frame. The algorithm allows for a temporary "zombie" state in order to cope with the problem of missing peaks that have not been detected in one frame between two other frames that contain candidate matching peaks.

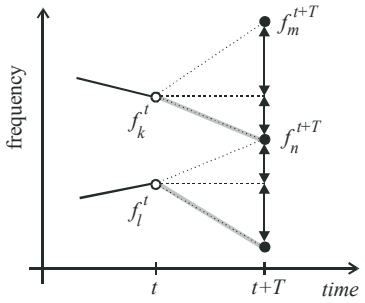

Fig. 3. MQ tracking algorithm: seeking for best match upon minimum absolute difference of frequency. Final decisions (shown in grey) due to the preference of closest distance

The main advantages of this algorithm are its simplicity and low delay resulting from taking into account only the local differences between two consecutive frames. The main drawbacks result from the facts that absolute frequency difference is calculated instead of a relative difference, and that amplitudes of spectral peaks are not taken into account. Consequently, their algorithm fails at tracking of signals with deep pitch modulations leaving many trajectories broken

(cf fig. 4) and also it attempts at connection of spurious peaks having low amplitude to peaks in the main lobe.

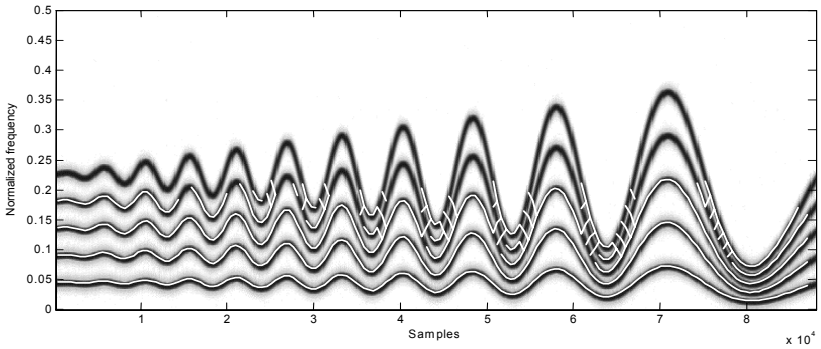

Fig. 4. An example of unsuccessful tracking of the MQ algorithm in the case of deep frequency modulation (synthetic signal).

Depalle et al. developed an advanced partial tracking technique using a Hidden Markov Model [6]. They defined a heuristic cost function for inclusion of given peak into the trajectory. This function relies on state transition probabilities dependent on amplitude and frequency deviations between the two last inserted peaks and the peak candidate. In order to globally optimize all trajectories, dynamic programming is applied for all partials and all frames. This algorithm is very innovative, as it handles crossing between trajectories and allows to avoid inclusion of noisy peaks into the trajectories, however it does not cope with missing peaks. The main drawback of this approach is the prohibitive computational cost related to global optimization. Furthermore, this algorithm is immanently unable to operate in real time.

Lagrange et al. proposed an application of adaptive linear prediction for predicting the evolutions of partial frequencies in time [5,7]. Their enhanced partial tracking algorithm relies on the Burg method for calculating the predictor coefficients in the AR model. A number (2-8) of previous peaks from the already tracked partial is involved in estimation of the predicted position of the current peak. Matching of peaks depends on the absolute value of the prediction error which should be less than assumed threshold. This algorithm is able to properly track partials with deep frequency variations. The main drawback of this approach is that the assumptions on the predictive behavior of trajectories are often too restrictive due to the fact that reliable prediction involves a significant analysis lag, while speech and music are often highly nonstationary.

4. PROPOSED TECHNIQUE FOR PARTIAL TRACKING

We propose a simple extension of the classical partial tracking algorithm by the inclusion of additional peak information. Application of a first-order AM-FM model in the spectral analysis allows for estimation of the amplitude and frequency slope (modulation rate) according to the linear chirp model (3).

$$x(t) \cong \sum_{k=1}^{K} A_k(t) \exp(\alpha_k t) \cos\left[\varphi_k + 2\pi \left(f_k t + \beta_k t^2\right)\right]. \tag{3}$$

An accurate solution of the estimation of α_k and β_k parameters alongside the f_k and A_k is based on the first and second derivatives of the magnitude and phase spectra around each magnitude peak, as proposed by Abe and Smith [6]. The experimental results show that the estimators are sufficiently robust against noise in order to support more accurate tracking.

We model each trajectory as a sequence of linear segments according to the estimated value of candidate peaks and corresponding slope factors. The likelihood of track connection between peaks of peaks S_k and S_m is estimated using a simple two-way mismatch measure (4),

$$L_{S_k^t, S_m^{t+T}}\{S_k^t \mapsto S_m^{t+T}\} = \frac{1}{1+\mu\left(S_k^t, S_m^{t+T}\right)}. \tag{4}$$

where the mismatch measure $\mu(S_k,S_m)$ is the area of the trapezoid between corresponding peaks normalized by the geometric mean of peak values (fig. 5). This normalization accounts for relative frequency deviation thus making equal chances that partials related to higher harmonics are properly tracked.

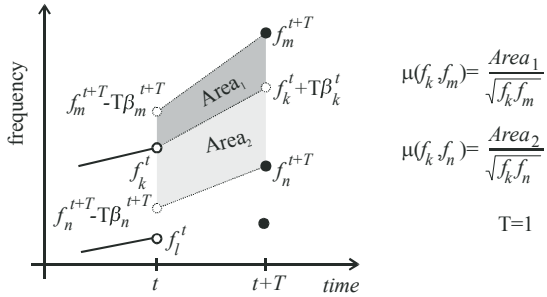

Fig. 5. Peak tracking according to the mismatch measure: the wining peak f_m is selected as best matching to f_k due to the observation that $\mu(f_k,f_m) < \mu(f_k,f_n)$, even though $|f_k - f_m| > |f_k - f_n|$.

In the proposed technique, the likelihood is calculated independently due to frequency (L_f) and amplitude (L_A) mismatch, using the respective estimated slopes, α_k and β_k. Each trajectory is extended along that peak, whose joint likelihood is the highest, provided that both L_A and L_f are greater than 0.5.

5. QUALITY ASSESSMENT FOR TRACKING ALGORITHMS

In order to compare the performance of various tracking algorithms we propose a special quality assessment methodology. The main problem here is that different operating conditions of the modeling algorithm (various levels of difficulty due to the character of the signal as well as due to the amount of disturbing noise) impose both partial estimation errors and tracking problems. In order to assess only the quality of tracking, a conditional error rate should be measured.

The methodology is based on comparing the tracking results with ground truth – known set of trajectories that are perfectly characterizing given test signal. Such data are easily obtained with artificially prepared test signals. Having the reference trajectories known, it is possible to measure the degree of conformance for the algorithm under test. For this, we propose to check every peak belonging to all partial trajectories by seeking for nearest peak in the reference data. We count a tracking error each time a peak is wrongly connected to a different peak whose g.t. counterpart belongs to a different trajectory, as well as each time when there is no connection while there should be one. The ratio of tracking errors to the total amount of proper and erroneous trajectory segments defines our error rate [%]. This measure is examined as a function of the amount of disturbing noise.

6. SIMULATION RESULTS

In our experiments we generate quasi-periodic signal with a specific number of harmonics and deep frequency modulations (as shown on the underlying spectrogram in fig. 4). The signal is disturbed by AWGN at various SNR settings. Figure 6 shows the determined number of correctly tracked peaks as a function of the amount of noise for the MQ algorithm and the new proposed technique.

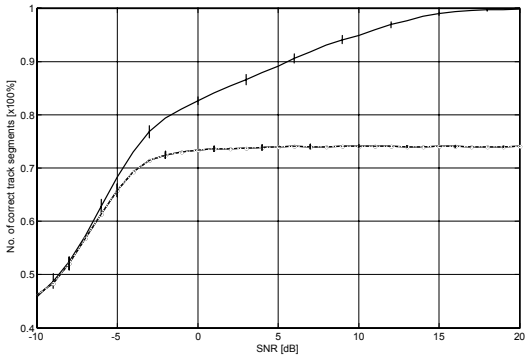

Fig. 6. Performance of the MQ algorithm (dashed) versus the proposed tracking technique (solid) with various amounts of additive white Gaussian noise.

Simulation results show that for sufficiently high signal to noise ratio allowing for a robust detection and estimation of sinusoidal parameters (SNR>0dB), our partial tracking clearly outperforms the MQ algorithm. At higher SNR levels, the proposed algorithm offers a 100% tracking accuracy as compared to the ground truth, while it is never possible with the classical algorithm.

7. CONCLUSIONS

A simple, but very efficient partial tracking technique is proposed for application to sinusoidal modeling. The technique is robust against white noise and is able to almost perfectly track sinusoidal partials with deep frequency modulations provided the amount of noise does not significantly decrease the accuracy of parameter estimation. The important advantage of our technique is its negligible computational complexity and operation without lag, which makes it perfect for applications in online signal processing and coding.

REFERENCES

[1] McAulay R.J., Quatieri T.F.: Speech Analysis/Synthesis Based on Sinusoidal Representation, IEEE Trans. Acoustics, Speech and Signal Processing, Vol. 34, No.4, August 1986, pp. 744-754.

[2] Smith J.O., PARSHL: A Program for the Analysis/Synthesis of Inharmonic Sounds Based on a Sinusoidal Representation, International Computer Music Conference, ICMC'87, Tokyo, 1987.

[3] Serra X., Smith J.O., Spectral modeling synthesis: A Sound Analysis/Synthesis System Based on a Deterministic Plus Stochastic Decomposition, Computer Music Journal, vol. 14, no. 4, Winter 1990, pp. 12–24.

[4] Marchand S., Keiler F.: Survey on Extraction of Sinusoids in Stationary Sounds, Digital Audio Effects (DAFx'02) Conference, September 2002, Hamburg, Germany, pp 51-58.

[5] Lagrange M., Marchand S., Rault J.-P.: Tracking Partials for the Sinusoidal Modeling of Polyphonic Sounds, IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP '05, March 2005, Philadelphia, USA, pp 229-232.

[6] Depalle P., Garcia G., Rodet X, Tracking of Partials for Additive Sound Synthesis using Hidden Markov Models, IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-93, Minneapolis, USA, April 1993, pp. 225-228.

[7] Lagrange M., Marchand S., Rault J.-P.: Using linear prediction to enhance the tracking of partials, IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP 2004, Quebec, Canada, May 2004, pp. 241-244.

[8] Abe M, Smith J.O., AM/FM Rate Estimation and Bias Correction for Time-Varying Sinusoidal Modeling, CCRMA Report No. STAN-M-118, Stanford University, October 2004.