

SAMPLE-BASED ENGINE NOISE SYNTHESIS USING A HARMONIC SYNCHRONOUS OVERLAP-AND-ADD METHOD

Jan JAGLA*

Julien MAILLARD *

Nadine MARTIN[†] Member, IEEE

* CSTB, Acoustic and Lighting Department, Paris-Est University,
24 rue Joseph Fourier, 38400 Saint Martin d'Hères, France

[†] GIPSA-lab, Grenoble Institute of Technology / CNRS
961 rue de la Houille Blanche, 38400 Saint Martin d'Hères, France

ABSTRACT

In order to synthesize the noise of internal combustion engines at arbitrary speed variations, an overlap-add-based method is proposed to be applied to sound samples extracted from a prerecorded engine noise at a continuously-varying speed. The objective of this new method is to minimize the phase discontinuities between a finite number of harmonics of the sample sounds likely to be concatenated. The harmonicity of the synthesized signal is then preserved which greatly enhances the acoustic quality of the restitution. Moreover, the synthesis stage is well suited for real time applications due to a low computational load compared to existing approaches. The proposed analysis-synthesis method can be applied to any harmonic sound whose spectral content depends solely on its fundamental frequency. By analogy to existing overlap-and-add methods, the method introduced here is referred to as Harmonic Synchronous Overlap and Add (HSOLA).

Index Terms— engine noise, signal synthesis, real time, overlap-and-add, harmonic signal

1. INTRODUCTION

The real time synthesis of engine noise has numerous applications in the field of road traffic auralization, but also in other areas such as driving simulators [1] or evaluation of exterior and interior vehicle sound quality [2, 3]. The existing methods are not optimal for real time applications. Therefore, a new method to synthesize engine sounds but also other types of non-stationary harmonic signals whose spectral content depends only on the instantaneous fundamental frequency value is introduced in the following. It requires very low computational load compared to existing methods. It is based on the principles of overlap-and-add (OLA) algorithms such as the Synchronous OLA (SOLA) introduced by Roucos [4] and discussed by Jones [5], the Pitch Synchronous OLA (PSOLA) [6] and the Waveform Similarity OLA (WSOLA) [7]. These methods were originally developed for pitch-scale and time-scale modifications of speech. The method introduced here intends to create a dataset of samples ordered with respect to their fundamental frequency varying according to the speed variations of the engine. The dataset is constructed such that any continuous evolution of the fundamental frequency can be rendered by overlapping and adding properly selected sound samples. The prevailing improvement of this method over existing OLA algorithms is the phase continuity of the harmonics of the synthesized signal. In fact, the pitch synchronization or the waveform similarity synchronization used in the PSOLA and WSOLA algorithms is not sufficient to preserve the harmonicity in

the synthesized signal when applied to engine noise. The method introduced here is therefore referred to as Harmonic Synchronous Overlap and Add (HSOLA).

2. CONSIDERED ENGINE NOISE MODEL

An engine noise sound signal is essentially harmonic due to the sequential explosions occurring in the engine cylinders. In this paper, only the four-stroke internal combustion engine is considered as it is used for most production cars, trucks and motor bikes. In this case, the explosion frequency is given by $F_{ex} = N_c \text{ rpm}/120$, where N_c is the number of cylinders in the engine and rpm the number of revolutions per minute. Two subsequent explosions do not occur in the same cylinder, thus, they are not identical. Likewise, two subsequent engine revolutions do not contain the explosions of the same cylinders. On that account, two other frequencies and their harmonics enrich the spectrum of an engine sound, $F_{en} = \text{rpm}/60$ and $F_c = \text{rpm}/120$, where F_{en} is the frequency corresponding to an engine revolution and F_c is the fundamental frequency corresponding to a complete cycle of the engine (two revolutions). In this paper, the noise emitted by an internal combustion engine is supposed to be fully characterized by its fundamental frequency and thus no dependency on the load applied to the engine is considered. An engine noise at constant speed can thus be modeled as

$$x_{F_0}[n] = \sum_k A_{F_0,k} \sin \left(2\pi \frac{kF_0}{F_s} n + \phi_{F_0,k} \right) + N_{F_0}[n], \quad (1)$$

where n is the discrete time index, $F_0 = F_c = \text{rpm}/120$ is the fundamental frequency, $A_{F_0,k}$ and $\phi_{F_0,k}$ are respectively the amplitudes and the initial phases of the harmonics and $N_{F_0}[n]$ is the stochastic component of the engine noise. The amplitudes, phases and spectral coloration of the stochastic noise depend only on the fundamental frequency F_0 . In the following, the order of the most energetic component is noted k_{Amax} . In practice, for a N_c -cylinder engine $k_{Amax} = N_c$.

3. OVERVIEW OF THE HSOLA APPROACH

3.1. Input of the HSOLA Approach

The input signal to the HSOLA method can be obtained by placing a microphone near the engine compartment and running the engine with no gear engaged over the available rpm range. The corresponding fundamental frequency range is noted $[F_0^{min}; F_0^{max}]$. The variation of the engine speed (and thus of the fundamental frequency F_0)

must be slow enough to assume its local stationarity on at least one period duration. The input signal with continuously varying engine speed can thus be modeled locally by equation 1.

3.2. Analysis and sample dataset creation

The analysis stage aims at iteratively collecting sound samples from the input signal at equally spaced fundamental frequencies $F_0^i = F_0^{min} + i(F_0^{max} - F_0^{min})/(M - 1)$, $i = 0..M - 1$, where M is the total number of sound samples to extract, generally set to about one thousand for common engines of most production cars, trucks and motorbikes and i is the iteration index.

First, the evolution of the fundamental frequency of the input signal must be estimated. A variety of robust methods for fundamental frequency tracking exist in the literature. In this application, a simplified version of the YAAPT method introduced by Zahorian [8] is used. Then, for every fundamental frequency F_0^i for which a sound sample has to be extracted, a search interval I_i is defined as the time segment of the input signal for which the value of the frequency $k_{Amax}F_0^i$ is between $k_{Amax}F_0^i - \Delta F$ and $k_{Amax}F_0^i + \Delta F$ (where ΔF can be set to $(F_0^{max} - F_0^{min})/M$ so that the $\{I_i\}_{i=0..M-1}$ intervals are disjoint). A sound sample is then extracted with an appropriate temporal window from each interval I_i . The fundamental frequency is assumed to be constant on the length of each extracted sound sample so that it can be modeled by equation 1. The main purpose of the HSOLA analysis will be to determine the optimal extraction instant, referred to as a pitch mark by analogy to the PSOLA terminology. Section 4 will describe this approach.

3.3. Synthesis

Once the dataset of sound samples is created, any arbitrary evolution of the fundamental frequency can be synthesized by applying the overlap-and-add algorithm on properly selected sound samples. The sample selection and the overlap-and-add constitute the synthesis stage of the HSOLA method, where the phase continuity is preserved for the harmonics, as described in Section 5.

4. ANALYSIS AND DATASET CREATION

4.1. Phase match optimization

To locate the optimal extraction instant in each interval I_i (defined in Section 3.2), a set of potential pitch marks is defined as the discrete time indexes $\{n_{i,l}\}_{l=1..l_{i,max}}$ (where l is the index of the potential pitch mark in interval I_i) where the initial phase of the harmonic component $k_{Amax}F_0^i$ is null for the portion of the input signal starting at instant $n_{i,l}$ (modeled by $x_{F_0^i}[n - n_{i,l}]$ according to equation 1). These potential pitch marks are obtained by calculating the cross-correlation function between a sinusoid of frequency $k_{Amax}F_0^i$ and the portion of the input signal defined by the interval I_i . The sinusoid starts at zero phase and its length is set to the length of the sample to be extracted. Thus, the potential pitch marks correspond to the local maxima of the calculated cross-correlation.

As a first approach referred to as single harmonic phase match, the optimal pitch mark for frequency F_0^i is obtained as the global maximum of the cross-correlation function.

As a second approach referred to as multiple harmonic phase match, the phases of higher order harmonics are taken into account to further improve the harmonicity of the synthesized signal. In the following, the initial phase of the frequency component kF_0^i of the portion of the input signal starting at instant $n_{i,l}$ is noted $\phi_{F_0^i,k,l}$. The

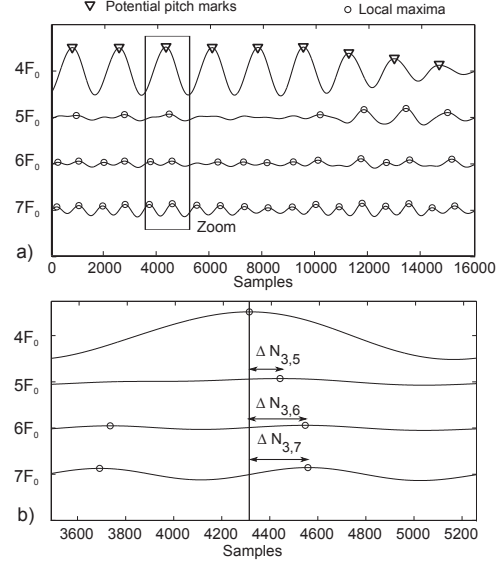


Fig. 1. a) Cross correlation functions between the segment I_i of a four-cylinder engine noise signal and sinusoids of frequencies $4F_0 = k_{Amax}F_0$ to $7F_0$, b) Zoom in the region in the frame corresponding to the third potential pitch mark ($l = 3$)

phases $\phi_{F_0^i,k,l}$ of a finite number of harmonics kF_0^i ($k > k_{Amax}$) are estimated for all the potential pitch marks $\{n_{i,l}\}$ and compared to the phases of sound samples with close fundamental frequencies extracted in the preceding iterations of the algorithm. The phases are estimated from the cross-correlations between the signal segment I_i and sinusoids of frequencies kF_0^i . See Figure 1a. for a four-cylinder engine noise case. The phases $\phi_{F_0^i,k,l}$ are

$$\phi_{F_0^i,k,l} = 2\pi \Delta N_{l,k} \frac{kF_0^i}{F_s}, \quad (2)$$

where k is the order of the considered harmonic, $\Delta N_{l,k}$ is the difference in samples between the potential pitch mark $n_{i,l}$ and the closest local maximum of the cross-correlation function calculated for the k^{th} harmonic (see Figure 1b) and F_s is the sampling frequency.

To choose the optimal pitch mark $n_{i,l,opt}$ among the set of potential pitch marks $\{n_{i,l}\}$, the mean phase of each harmonic over a finite number N_s of previously extracted samples is computed as

$$\overline{\phi_{F_0^i,k}} = \frac{\sum_{j=1}^{N_s} \phi_{F_0^i,k,l_{i,opt}}}{N_s}, \quad (3)$$

where $\phi_{F_0^i,k,l_{i,opt}}$ is the phase of the k^{th} harmonic of the extracted sound sample for frequency F_0^{i-j} during previous iterations.

Then, for each potential pitch mark, a phase matching criterion $C(l)$ is defined as the sum over the considered higher order harmonics of the phase differences $(\overline{\phi_{F_0^i,k}} - \phi_{F_0^i,k,l})$ weighted by the magnitudes of the cross correlations' local maxima in A-weighted decibels such that

$$C(l) = \sum_k \left| \overline{\phi_{F_0^i,k}} - \phi_{F_0^i,k,l} \right| A_{i,l,k}, \quad (4)$$

where $A_{i,l,k}$ is the magnitude in db(A) of the local maximum of the cross-correlation function used to calculate $\phi_{F_0^i,k,l}$.

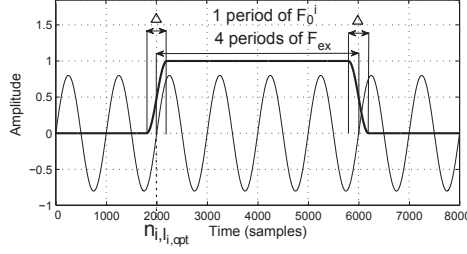


Fig. 2. Tukey window (bold line) applied to the extraction of a sound sample (schematic representation) corresponding to a fundamental frequency $F_0^i = F_{ex}/4$ (four cylinder engine). The $n_{i,l,opt}$ position is the selected pitch mark for the considered fundamental frequency.

The criterion $C(l)$ estimates the difference between the phase of the harmonics of the sound sample corresponding to the potential pitch mark $n_{i,l}$ for frequency F_0^i and the phase of the same harmonics of the N_s sound samples extracted for frequencies F_{0i-N_s} to F_{0i-1} . The optimal pitch mark $n_{i,l,opt}$ minimizes $C(l)$.

Experiments were conducted to determine practical values for the number of harmonics and the number of extracted samples N_s to consider for the calculation of $\phi_{F_0^i,k}$. Good results were obtained using about 8 harmonics. In practice, higher numbers of harmonics increase the computational load of the algorithm without improving the quality of the synthesis. In fact, correlations exist between the phases of the harmonics of an engine noise and hence, synchronizing the phases of low order harmonics also improves the phase matching of higher order harmonics. The number N_s is chosen such that the criterion $C(l)$ guarantees the phase continuity between sound samples that are likely to be concatenated during the synthesis stage. In practice, N_s can be set to 10 as explained in Section 5.2.

4.2. Creation of the sample dataset

Once a pitch mark has been optimally placed, the corresponding sound sample is extracted. The standard choice in existing OLA methods for the sample extraction envelope is a Hanning window containing an integer number of pitch periods. This window requires a 50% overlap during the synthesis stage and thus is not optimal in terms of complexity. A shorter overlap window such as in Figure 2 with a Tukey window is preferable.

The window length is adjusted to contain an integer number of pitch periods plus Δ (half the length of the attack ($\Delta/2$) and half the length of the release ($\Delta/2$), see Figure 2). During the synthesis stage, successive sound samples overlap by only Δ samples. To further simplify the algorithm, the length Δ is kept constant and set to 128 samples in practice, i.e., approximately 3 ms for 44.1 kHz sampled signals. In the particular case of engine noise, the extraction window is set to contain one period of the fundamental frequency plus Δ . This choice is relevant since a single period of the fundamental comprises the explosions of all cylinders of the engine. Hence, each extracted sound sample contains all the information about the engine sound at a given rotational speed.

5. SYNTHESIS

5.1. Overlap-and-add processing

The sample dataset created during the analysis stage allows the synthesis of any fundamental frequency evolution by overlapping and

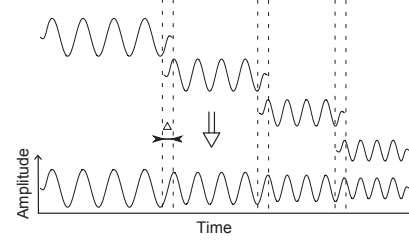


Fig. 3. Phase matching using the HSOLA synthesis scheme applied to varying fundamental frequency sound samples.

adding the sound samples. The sound samples are selected (see Section 5.2) and concatenated with an overlap of Δ samples for the sample envelope defined in the previous section (see Figure 3).

The selected overlap is sufficient to suppress the perceptive artifacts caused by slight differences in frequency and amplitude between successive sound samples. The pitch and harmonic synchronization performed in the analysis stage significantly improves the quality of this synthesis scheme with low overlap since the phases at the samples boundaries are tuned to be continuous.

5.2. Sample selection

A specific sample selection algorithm is implemented to ensure high quality synthesis. In the case of constant or slowly varying fundamental frequencies (engine regimes), a pseudo-random oscillation is added to the selected sample index in the dataset. The audible artifacts due to the successive selection of the same sample are thereby avoided. This oscillation is approximately zero mean to ensure that the average fundamental of the synthesized signal matches the target value. The maximum amplitude α of the oscillation must be large enough to provide sufficient choice for the selection of sound samples but should not induce large fundamental frequency discontinuities. In practice, $\alpha = 5$ represents a good trade-off.

The value α also affects the choice of the number of sound samples N_s to consider in the calculation of $\phi_{F_0^i,k}$ during the analysis stage (see Section 4.1). To ensure a good phase continuity between sound samples likely to be concatenated, N_s should be set to the highest possible variation of the sample index between successively selected sound samples. Fast evolutions of the engine speed seldom occur. Hence, N_s can be set to 2α considering the pseudo-random oscillation as the highest possible variation of the sample index.

The main advantage of the proposed HSOLA analysis/synthesis scheme is that most of the algorithmic complexity is transferred to the analysis stage which is performed offline in the pre-processing stage. The operations during the synthesis stage are limited to the selection of each sound sample in the dataset and the addition of Δ samples per selected sound sample. Therefore, the HSOLA method is particularly well suited to the real time synthesis of engine sounds.

6. PERFORMANCE OF THE HSOLA APPROACH

6.1. Spectrogram based evaluation

The analysis of the spectral content of a synthesized signal with the proposed HSOLA method provides valuable material to assess the performance of the algorithm. Two engine noise signals were recorded with different variations of the fundamental frequency. The first signal, referred to as *input 1*, is used for the creation of two

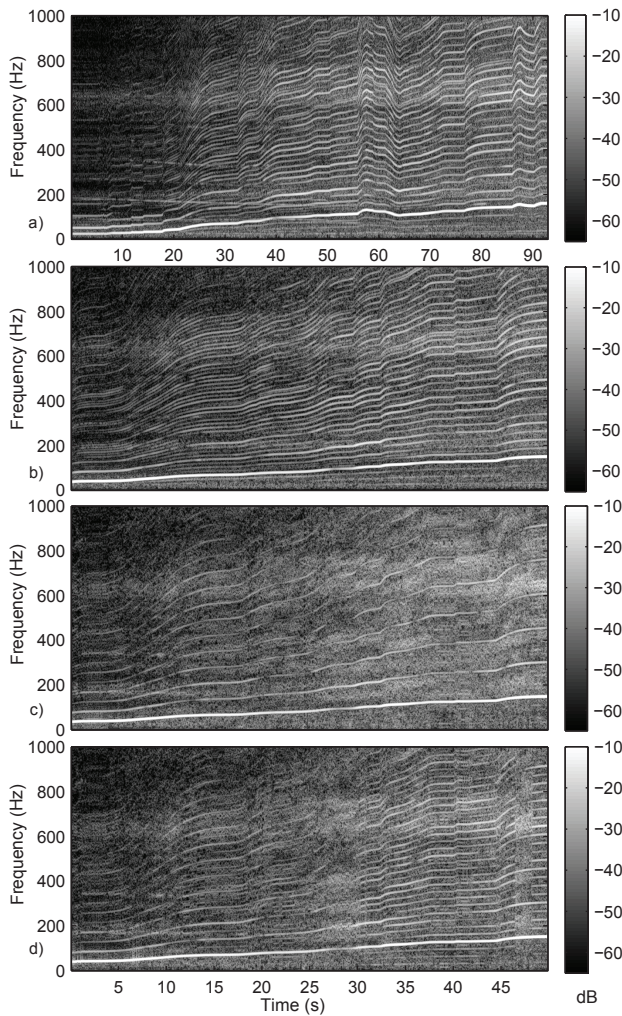


Fig. 4. Spectrograms of *input 1* (a), *input 2* (b), synthesized output of *input 2* speed variation with single harmonic (c) and with multiple harmonic (d) phase match.

sample-datasets, first one using pitch marks based on single harmonic phase match and second one using multiple harmonic phase match as defined in Section 4.1. The fundamental frequency of the second signal (referred to as *input 2*) is used as the command to synthesize engine noise signals.

Figure 4 shows the spectrograms of the two input signals and the two synthesized ones. They are obtained using Hanning windows of length 200 ms zeropadded to 743 ms (32768 points) with a 50 % overlap. The sampling frequency is 44100 Hz. To better display the harmonic content of the synthesized signals, only a zoom between 0 and 1000 Hz is shown.

The difference between spectrograms Fig.4(c) and Fig.4(d) illustrates the importance of the phase continuity between sound samples on the harmonicity of the synthesized signal. With single harmonic phase match, Fig.4(c), only the explosion frequency F_{ex} and its harmonics are rendered properly. On the other hand, multiple harmonic phase match, Fig.4(d), succeeds in preserving most of the harmonic content of engine noise signals. The phase matching of multiple harmonics greatly improves the quality of the synthesized signal.

6.2. Perceptive evaluation

In terms of sound quality, informal listening tests show that the sound signal synthesized with the HSOLA method is perceptually close to the *input 2* signal. The tests also show clear improvement over existing approaches. Moreover, in the case of stationary fundamental frequency commands, the random sample selection algorithm presented in 5.2 is effective. There are no audible sample repetitions and the resulting perceived pitch sounds stationary.

More thorough listening experiments based on psychoacoustic test procedures will be conducted to assess the performance of the different steps of the algorithm such as the sample location process and the random sample selection algorithm.

7. CONCLUSION

In the present paper, a method to synthesize engine noise has been introduced. This method termed HSOLA can be applied to synthesize any harmonic signal with varying fundamental frequency if its spectral content depends only on the instantaneous fundamental frequency value. It aims at extracting from a recorded internal combustion engine noise, a dataset of sound samples corresponding to equally sampled fundamental frequency values. The sound samples can be concatenated using an appropriate sample selection and overlap-and-add algorithm to simulate any fundamental frequency evolution. The prevailing feature of the HSOLA method is the optimization of the sample location in the input signal so that the phases of the harmonics of the synthesized signal tend to be continuous. It is shown that this continuity ensures that the harmonicity of engine noise is preserved. Moreover, the low complexity of the synthesis of engine signals with the HSOLA method makes it particularly relevant for real time synthesis applications.

8. REFERENCES

- [1] J. Van Rensburg, "Phase vocoder technology for the simulation of engine sound," *Int. Journal of Modern Physics C*, vol. 17, pp. 721–731, 2006.
- [2] K. Genuit and W. R. Bray, "Prediction of sound and vibration in a virtual automobile," *Journal of Sound and Vibration*, vol. 36, no. 7, pp. 12–19, 2002.
- [3] J-F. Sciabica, M-C. Bézat, V. Roussarie, R. Kronland-Martinet, and S. Ystad, "Towards timbre modeling of sounds inside accelerating cars," *Auditory Display*, vol. 5954, pp. 377–391, 2009.
- [4] S. Roucos and A. Wilgus, "High quality time-scale modifications for speech," *ICASSP'85*, pp. 236–239, 1985.
- [5] D. L. Jones and T. W. Parks, "Generation and combination of grains for music synthesis," *Computer Music Journal*, vol. 12, pp. 27–34, 1988.
- [6] E. Moulines and F. Charpentier, "Pitch synchronous waveform processing techniques for text-to speech synthesis using diphones," *Speech Communication*, vol. 9(5/6), pp. 453–467, 1990.
- [7] W. Verhelst and M. Roelands, "An overlap-add technique based on waveform similarity (wsola) for high quality time-scale modifications of speech," *ICASSP'93*, pp. 554–557, 1993.
- [8] S. Zahorian and H. Hu, "A spectral/temporal method for robust fundamental frequency tracking," *Journal of the Acoustical Society of America*, vol. 123, pp. 4559–4571, 2008.