HAL
open science

The expected amplitude of overlapping partials of harmonic sounds

Chunghsin Yeh, Axel Roebel

To cite this version:

Chunghsin Yeh, Axel Roebel. The expected amplitude of overlapping partials of harmonic sounds. International Conference on Acoustics, Speech and Signal Processing (ICASSP'09), 2009, Taipei, Taiwan. pp.3169-3172. hal-01161312

HAL Id: hal-01161312

https://hal.science/hal-01161312

Submitted on 8 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THE EXPECTED AMPLITUDE OF OVERLAPPING PARTIALS OF HARMONIC SOUNDS

Chunghsin Yeh and Axel Roebel

IRCAM/CNRS-STMS Analysis/Synthesis team
1 place Igor Stravinsky 75004 Paris France

ABSTRACT

In analyzing polyphonic signals, the handling of overlapping partials is one important problem. The assumptions usually made for partial overlaps are the additivity of the linear spectrum or that of the power spectrum. In this study, the expected amplitude of two overlapping partials is derived based on the assumption that the partials overlap at the same frequency and the phase is uniformly distributed. An overlap chain rule algorithm is proposed to estimate the amplitude for the case that more than two partials overlap. The proposed algorithm has demonstrated its better accuracy over the usual two model assumptions.

***Index Terms*—** Overlapping partials, multiple-F0 estimation, source separation

1. INTRODUCTION

In the problem of multiple-F0 (fundamental frequency) estimation and blind source separation, the additive signal model is usually used

$$y[n] = \sum_{m=1}^{M} y_m[n] + z[n] \tag{1}$$

where n is the discrete time index, M is the number of harmonic sources (the polyphony), $y_m[n]$ is the quasi-periodic part of the mth source and $z[n]$ is the residual part. There are three fundamental model assumptions involved: the residual model $z[n]$, the source model $y_m[n]$ and the source interaction model $\sum_{m=1}^{M} y_m[n]$. Very often, the source models are learned beforehand and the problem is to decompose the observed spectrum into a combination of the spectral models of sound sources. For example, several Non-negative Matrix Factorization (NMF) methods decompose the linear spectrum [1] [2], while the harmonic temporal structured clustering (HTC) method [3] and the specmurt method [4] decompose the power spectrum. Little attention is drawn to the treatment of overlapping partials, which is closely related to the source interaction model. Two assumptions are usually made implicitly for overlapping partials: the *additivity of linear spectrum* and the *additivity of power spectrum*. In this article, the expected amplitude of overlapping partials is studied, taking into account the statistical properties of phase. In Section 2, the point of view of this study is first described, followed by the mathematical derivation of the expected amplitude of two overlapping partials. In the case that more than two partials overlap, an algorithm based on the chain rule is further proposed. To compare the proposed algorithm with the usually made models: additivity of linear or power spectrum, an evaluation framework is proposed in Section 3. Finally, discussions and conclusions are given.

2. MATHEMATICAL TREATMENT OF OVERLAPPING SINUSOIDS

To facilitate the study of overlapping partials, it is assumed that the partials are represented by sinusoids of the same frequency but of different phases. Considering K sinusoids overlap, the resulting sinusoid can be written as [5]

$$A\cos(\omega n + \phi) = \sum_{k=1}^{K} A_k \cos(\omega n + \phi_k) \tag{2}$$

where each sinusoid has amplitude A_k, phase ϕ_k and frequency ω. Expand the above equation using trigonometric identity and obtain

$$A = \sqrt{\Big[\sum_{k=1}^{K} A_k \cos(\phi_k)\Big]^2 + \Big[\sum_{k=1}^{K} A_k \sin(\phi_k)\Big]^2} \tag{3}$$

The estimation of the resulting amplitude A of overlapping partials is related to the decomposition stage in the problem of multiple-F0 estimation or blind source separation: given the hypothetical sources, namely the F0s and the related spectral envelope models, the scaling factors of source models are to be estimated such that the resulting combination matches the observed spectrum. Some methods optimize the decomposition by estimating the amplitudes and phases of each partials [6] [7]. In the following, the expected amplitude of overlapping partials is proposed, which does not require the estimation of the related phases. It is assumed that the amplitudes of all partials are known, or can be represented by the spectral envelopes of source models.

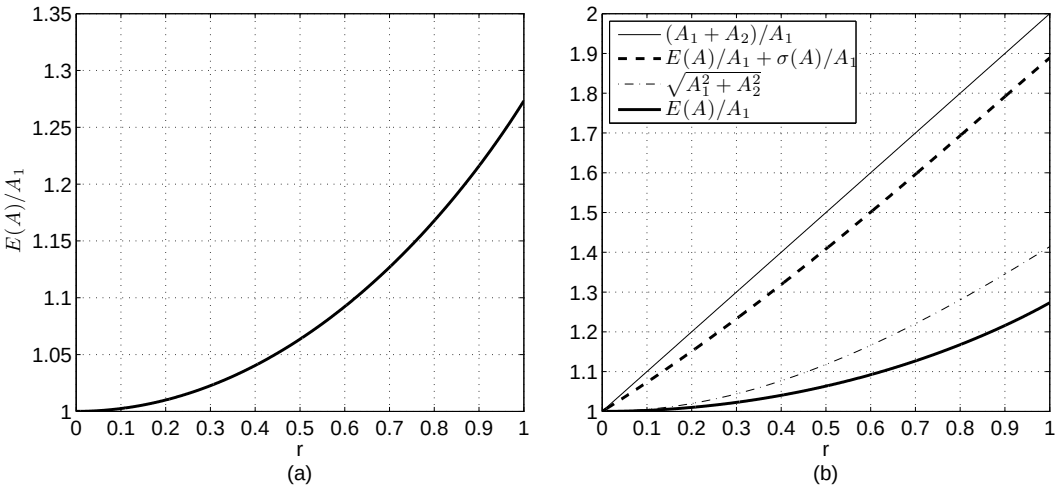

Fig. 1. (a) The expected amplitude as a function of the amplitude ratio between two partials; (b) Comparison between the expected amplitudes and the other two overlap models.

2.1. Expected amplitude of two overlapping partials

The overlap of two partials can be represented by summing two sinusoids of the same frequency but of different amplitudes and a phase difference:

$$s = s_1 + s_2 = A_1 \cos(\omega_1 n + \phi) + A_2 \cos(\omega_1 n) \quad (4)$$

In the case that the frequencies of the partials are very close but not exactly the same, the frequency difference can be also modeled by the phase difference term. The amplitude of s is then

$$A = \sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos\phi} \quad (5)$$

It is assumed that the phase difference ϕ is uniformly distributed between $-\pi$ and π, i.e., its probability density function is $f(\phi) = 1/2\pi$ within the related interval and $f(\phi) = 0$ elsewhere. The expected value of A can thus be calculated:

$$\begin{aligned} E(A) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos\phi}\,d\phi \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sqrt{A_1^2 + A_2^2 + 2A_1A_2(1-2\sin^2\frac{\phi}{2})}\,d\phi \\ &= \frac{A_1+A_2}{2\pi}\int_{-\pi}^{\pi}\sqrt{1-\frac{4A_1A_2}{(A_1+A_2)^2}\sin^2\frac{\phi}{2}}\,d\phi \\ &= \frac{A_1+A_2}{\pi}\int_{-\pi/2}^{\pi/2}\sqrt{1-\frac{4A_1A_2}{(A_1+A_2)^2}\sin^2\theta}\,d\theta \\ &= \frac{2(A_1+A_2)}{\pi}\int_{0}^{\pi/2}\sqrt{1-\frac{4A_1A_2}{(A_1+A_2)^2}\sin^2\theta}\,d\theta \\ &= \frac{2(A_1+A_2)}{\pi}E_p\left(\frac{2\sqrt{A_1A_2}}{A_1+A_2}\right) \end{aligned} \quad (6)$$

where

$$E_p(\mathrm{k}) = \int_0^{\pi/2}\sqrt{1-\mathrm{k}^2 sin^2\theta}\,d\theta \quad (7)$$

is the *complete elliptic integral of the second kind*. The variance of A is:

$$\begin{aligned} \mathrm{var}(A) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}(\sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos\phi} - E(A))^2 d\phi \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}(A_1^2 + A_2^2 + 2A_1A_2\cos\phi)d\phi + E(A)^2 \\ &- 2E(A)\frac{1}{2\pi}\int_{-\pi}^{\pi}\sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos\phi}\,d\phi \\ &= A_1^2 + A_2^2 - E(A)^2 \\ &= E(A^2) - E(A)^2 \end{aligned} \quad (8)$$

where $E(A^2)$ is the expected power. Without loss of generality, substitute $A_2 = rA_1 \le A_1$ (where $0 \le r \le 1$) into eq.(6) to have

$$\frac{E(A)}{A_1} = \frac{2(1+r)}{\pi}E_p\left(\frac{2\sqrt{r}}{1+r}\right) \quad (9)$$

This equation shows that as long as the amplitude ratio between two sinusoids is known, the increment of the overlapping amplitude relative to the stronger sinusoid can be deduced easily (see Figure 1 (a)). A source interaction model, or overlap model, for two overlapping partials, called the *expected amplitude model*, is thus defined. The standard deviation appears to be rather large (compare the thick solid line and the thick dash line in Figure 1 (b)), which implies the large uncertainty in the estimation of the amplitude of overlapping partials. The two assumptions usually made are

shown together in Figure 1 (b). The additivity of linear spectrum ($A_1 + A_2$) implies the maximum of the resulting amplitude that occurs when two sinusoids are *in phase*, that is, $\cos(\phi) = 1$. The additivity of power spectrum ($\sqrt{A_1^2 + A_2^2}$) in fact makes use of the expected power $E(A^2)$. This assumption is considered correct when the decomposition stage remains in the power domain [3] [4], but less appropriate when the linear amplitude of the partials is to be estimated, which is often required for the separation or synthesis stage.

2.2. The amplitude estimation of more than two overlapping partials

For the general case that more than two partials overlap, it is assumed that the expected amplitude can be approximated pair by pair using the proposed model. Given the amplitudes of all partials that overlap $\{A_1 > A_2 > \cdots > A_K\}$, the expected amplitude is consecutively estimated by using eq.(9). The order of pairing the overlapping partials is suggested from the largest to the smallest, which is meant to minimize the accumulated errors. The larger partials that tend to cause larger estimation errors are thus handled with a more appropriate estimation. Notice that this overlap chain rule is implied in the other two overlap models as well.

Algorithm 1: Overlap chain rule

input : The amplitudes of K overlapping partials $\{A_1 > A_2 > \cdots > A_K\}$
output: The estimated resulting amplitude A'

Initialization of $A' \leftarrow A_1$
for $k = 2$ *to* K **do**
  $A' \leftarrow A' \star A_k$ `/* apply the estimation for two partials */`
end
return A'

3. EVALUATION

In order to compare the expected amplitude model, the additivity model of linear spectrum and that of power spectrum, an evaluation framework is proposed. Following the description in the introduction of Section 2, a testing algorithm (see **Algorithm 2**) is designed to decompose the observed spectrum into a combination of source models which are assumed known in advance. The idea is to make use of the available source models such that the different methods for estimating the overlapping partial amplitudes can be evaluated. A previously proposed method for creating synthesized polyphonic music database facilitates the extraction of the true models of individual sound sources beforehand [8]. The synthesized music database allows not only the estimation of the true source models but also the precise inference of overlap positions where the overlap models are applied and evaluated.

Algorithm 2: Expected spectrum using overlap model

input : The observed signal of M mixing sources with fundamental frequencies $\mathrm{F0}_1 < \mathrm{F0}_2 < \cdots < \mathrm{F0}_M$ and known source models as partial amplitude sequences
output: The scaling factors $(c_m)_{m=1}^{M}$ for source models that minimize the modeling error

Initialization of c_1 by matching non-overlapped partials
for $m = 2$ *to* M **do**
  find non-overlapped partials: $\mathcal{U}$
  find partials overlapped with those of lower F0s but not those of higher F0s: $\mathcal{V}$
  do initialization of the expected amplitude of $\mathcal{V} \leftarrow$ **apply overlap model**
  if *exist* $\mathcal{U}$ **then**
    Initialization of c_m using $\mathcal{U}$ with LSE
  else
    Initialization of c_m using $\mathcal{V}$ with LSE
  end
  do optimize c_m using gradient descend for both $\mathcal{U}$ and $\mathcal{V} \leftarrow$ **apply overlap model**
end

Given the F0s and the spectral envelopes of the concurrent sound sources, the scaling factors are to be estimated such that an optimal decomposition of the observed spectrum is achieved. It is proposed to estimate the scaling factors in a progressive manner: from the source of the lowest F0, considered having the most non-overlapping partials, to that of the highest F0, considered having the most overlapping partials. To estimate the scaling factor c_m, the gradient descend method is used to optimize the estimation based on least square error (LSE) estimation:

$$\underset{c_m}{\operatorname{argmin}} |c_1 E_1(\omega_m) \star c_2 E_2(\omega_m) \star \cdots c_m E_m(\omega_m) - E(\omega_m)|^2 \tag{10}$$

where ω_m is the set of partial frequencies except for the partials that overlap with sources of higher F0s (scaling factors not yet estimated), E_m is the spectral envelope of the mth source and E is the observed spectral envelope at ω_m. The $\star$ stands for applying the overlap model. The optimal estimate of c_m makes use of non-overlapping partials and the partials that overlap with the sources from 1 to $m - 1$. The more accurate estimate of the scaling factors of source models should give rise to the smaller modeling error, which indicates a better overlap model assumption.

The evaluation is carried out for 26 pieces of synthesized music. In each STFT (short-time Fourier transform) analysis frame of 93ms, the overlapping partials are located if the source partials fall within the mainlobe of the corresponding peak in the spectrum of the polyphonic signal. That is,

the partials may not overlap with each other at the same frequency, which is the case of most real-world polyphonic signals. The modeling errors are evaluated for each analysis frame and the results are averaged for all testing pieces with respect to the number of overlapping partials (see Figure 2). For each number of overlapping partials, the modeling error is normalized by the sum of the related partial amplitudes.

The modeling error for non-overlapped partials (the number of overlap $= 1$) is about 2%, which demonstrates good estimation of the scaling factors by **Algorithm 2** and the trustworthiness of the evaluation setup. Of the three models, the expected amplitude model performs the best. The additivity of linear spectrum model has rather high errors compared to the other two models. In general, the additivity of power spectrum has similar performance to that of the expected amplitude model.

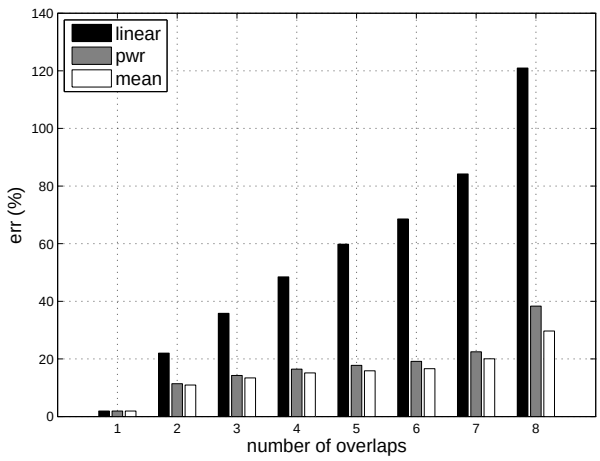

Fig. 2. Comparison of the three overlap models (linear: additivity of linear spectrum, pwr: additivity of power spectrum, mean: expected amplitude model) for the estimation of the amplitude of overlapping partials. The x-axis represents the number of partials that overlap and the y-axis indicates the modeling error rates.

4. CONCLUSIONS

A method for estimating the amplitude of overlapping partials is presented. It is based on the expected amplitude of two overlapping sinusoids. With the proposed overlap chain rule, the expected amplitude of more than two overlapping partials is estimated. The evaluation framework uses a synthesized polyphonic music database and a specially designed algorithm such that the effecting factors except the overlap model are minimized. Compared to the usual assumption made that the linear or power spectrum is additive, the proposed method demonstrates better accuracy for the number of overlaps up to eight. The additivity of linear spectrum, although being often implicitly used in quite a few previous studies, is rather worse. The advantage of the expected amplitude model is the handling of the phase information, taking into account its statistical distribution while avoiding the direct estimation of the related phases. This study provides an alternative overlap model that can be useful for the related research topics such as multiple-F0 estimation and source separation. For instance, the combination of hypothetical source models can make use of the overlap model to evaluate its likelihood with respect to the observed spectrum. Further improvements may focus on two aspects: the naive overlap chain rule causing the estimation error to accumulate, and taking into account the rather large variance of the expected amplitude model.

5. REFERENCES

[1] F. Sha and L. K. Saul, "Real-time pitch determination of one or more voices by nonnegative matrix factorization," *Advances in Neural Information Processing Systems*, vol. 17, pp. 1233–1240, 2005.

[2] A. Cont, "Realtime multiple pitch observation using sparse non-negative constraints," in *Proc. of the 7th Intl. Conf. on Music Information Retrieval (ISMIR 2006)*, 2006.

[3] H. Kameoka, T. Nishimoto, and S. Sagayama, "A multipitch analyzer based on harmonic temporal structured clustering," *IEEE Trans. on Audio, Speech and Language Processing*, pp. 982–994, 2007.

[4] S. Saito, H. Kameoka, K. Takahashi, T. Nishimoto, and S. Sagayama, "Specmurt analysis of polyphonic music signals," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 16, no. 3, pp. 639 – 650, March 2008.

[5] J. O. Smith, *Introduction to Digital Filters with Audio Applications*, W3K Publishing, 2007.

[6] Tuomas Virtanen, "Algorithm for the separation of harmonic sounds with time-frequency smoothness constraint," in *Proc. of the 6th Intl. Conf. on Digital Audio Effects (DAFx-03)*, London, UK, September 8-11 2003.

[7] John Woodruff, Yipeng Li, and DeLiangWang, "Resolving overlapping harmonics for monaural musical sound separation using pitch and common amplitude modulation," in *Proc. of 9th Intl. Conf. on Music Information Retrieval*, Philadelphia, USA, 2008.

[8] C. Yeh, N. Bogaards, and A. Roebel, "Synthesized polyphonic music database with verifiable ground truth for multiple f0 estimation," in *Proc. of the 8th Intl. Conf. on Music Information Retrieval (ISMIR'07)*, Vienna, Austria, 2007, pp. 393–398.