Physical principles driven joint evaluation of multiple F0 hypotheses

Chunghsin Yeh, Axel Roebel

Physical principles driven joint evaluation of multiple *F0* hypotheses

Chunghsin Yeh and Axel Röbel

IRCAM, Analysis-Synthesis team
1, Place Igor-Stravinsky 75004 Paris FRANCE
cyeh@ircam.fr and roebel@ircam.fr

Abstract

This article is concerned with the estimation of fundamental frequencies in polyphonic signals for the case when the number of sources is known. We propose a new method for joint evaluation of multiple *F0* hypotheses based on three physical principles: harmonicity, spectral smoothness and synchronous amplitude evolution within a single source, which are closely related to source segregation in auditory scene analysis. Given the observed spectrum a set of hypothetic partial sequences is derived and an optimal assignment of the observed peaks to the hypothetic sources and noise is performed. Hypothetic partial sequences are then evaluated by a new score function which formulates the guiding principles in a mathematical manner. The algorithm has been tested on a large collection of artificially mixed polyphonic samples and the encouraging results demonstrate the competitive performance of the proposed method.

1. Introduction

The estimation of the fundamental frequency, or *F0*, of a sound source from a given signal is an essential step for many signal processing applications. For the monophonic case there exist many approaches that achieve satisfying performance. Despite increasing research activities with respect to polyphonic signals the estimation of multiple *F0*s remains a challenging problem. In the following article, we propose a new method for multiple *F0* estimation under the assumption that the number of *F0*s is known in advance.

There exist several approaches for multiple *F0* estimation. A probabilistic signal modeling approach proposed in [3] applies specific prior distributions on the model parameters. This approach is computationally expensive and limited results are reported. In [16], a robust multipitch estimation is achieved by means of selecting reliable frequency channels as well as reliable peaks in the normalized correlograms. The technique has been reported to work for two-voice speech and the authors conclude that the proposed algorithm could be extended to more than two pitches. Klapuri's iterative multiple *F0* estimation algorithm handles most of the difficulties like estimating the number of *F0*s and treating the overlaps of coincident partials. Promising results are reported by evaluating a variety of polyphonic musical signals[10]. An iterative estimation and cancellation model has been proposed by de Cheveigné earlier in [4]. He compared an iterative approach and a full search approach which performs a joint evaluation. In this early study and later work in [5], he reported that a joint cancellation performs better than an iterative cancellation.

In fact, a joint evaluation strategy provides more flexibility in *F0* estimation for polyphonic signals. For a set of multiple *F0* hypotheses, spectral components in the interleaved spectrum could be reasonably allocated to each *F0* hypothesis and disturbed information provided by overlapped partials could be identified and taken care of in a more accurate way. Therefore, we propose a new method for jointly evaluating multiple *F0* hypotheses. Based on a generative spectral model, hypothetic partial sequences are constructed and evaluated using three general principles: harmonicity, smoothness of spectral envelopes and synchronous evolution of sinusoidal amplitudes. These physical principles are formulated as four criteria to construct the final score function which is used to rank the sets of *F0* hypotheses. The contribution of the following article consists, first, in a new proposition to make use of the hypothetic *F0*s to determine reliable information in the observed spectrum, and second, in a new mathematical interpretation of the guiding principles.

This paper is organized as follows. In section **2** the generative spectral model is described and three principles for quasiharmonic sounds are established. In section **3**, we introduce a frame-based *F0* estimation method using the proposed score function. In section **4**, experimental results are discussed which prove the competitive performance of the proposed method. Finally, further improvements are discussed and conclusions are drawn.

2. Generative quasiharmonic model

The proposed algorithm is based on a polyphonic quasiharmonic signal model of the following form

$$y[n] = \Big\{ \sum_{m=1}^{M} \sum_{h_m=1}^{H_m} a_{m,h_m}[n] \cos\big((1+\delta_{m,h_m}) h_m \omega_m n + \phi_m[n]\big)\Big\} + v[n], \quad (1)$$

where n is the discrete time index, M is the number of sources, H_m is the number of partials for the m-th source, ω_m represents the *F0* of source m and $\phi_m[n]$ denotes the

phase. In the current context these parameters are either fixed or of minor interest. The score function will make use of $a_{m,h}[n]$ and $\delta_{m,h}$, which are the time varying amplitude and the constant frequency detuning of the h-th partial, and also $v[n]$ which is the residual noise component. Generally it is supposed that the noise is sufficiently small such that a considerable part of the individual sinusoidal components can be identified.

Similar to [6] we understand the observed spectrum as generated by sinusoidal components and noise. Each spectral peak is characterized by its amplitude and frequency. A sinusoidal peak is assigned to one or more of the M sources in eq.(1), all unassigned peaks contribute to the noise component $v[n]$. The model supposes quasi-stationary frequency and, therefore, the sinusoidality of an observed peak is used to rate the requirement to include it into the quasiharmonic parts of the source model. Based on this model and given the observed spectrum and M, the most plausible *F0* hypotheses are going to be inferred.

To construct and evaluate hypothetic sources, we use three physical principles for quasiharmonic sounds stated in the following.

Principle 1: *Spectral match with low inharmonicity*. For a *F0* hypothesis, a hypothetic partial sequence HPS_{F0} is constructed by selecting harmonically matched peaks from the observed spectrum in such a way that $\delta_{m,h}$ are minimized. The set $\{HPS_{F0_m}\}_{m=1}^{M}$ should combinatorially "explain" the sinusoidal components in the observed spectrum. Under the assumption that the noise energy is small it is reasonable to favor *F0* hypotheses that explain more components of the observed spectrum as long as they are not contradicted by the following two principles.

Principle 2: *Spectral smoothness*. For natural quasiharmonic sounds, the spectral envelopes usually form smooth contours. While constructing HPS_{F0} of a source, the partials should be selected in a way such that $\{a_{m,h}\}_{h=1}^{H_m}$ results in a smooth spectral envelope. For partial sequences fitting well to **Principle 1**, those with smoother spectral envelopes are more probable to be originated from natural sources such as musical instruments.

Principle 3: *Synchronous amplitude evolution within a single source*. Partials belonging to the same source should have similar time evolution of the amplitudes $\{a_{m,h}\}_{h=1}^{H_m}$ collected in a *HPS*. If the partials of a hypothetic source match mostly to noisy peaks, they evolve in a random manner and thus do not have a synchronous amplitude evolution.

The three principles concerning the physical laws of quasiharmonic sounds are closely related to source segregation in auditory scene analysis. As discussed in [1], the human auditory system seems to segregate harmonically related spectral components forming a smooth envelope (p.232) and having a similar temporal evolution (p.575).

3. Multiple *F0* estimation

Based on the three principles described above, we design a multiple *F0* estimation system. The main task is to formulate these principles into four criteria serving as the core components in a score function for evaluating the plausibility of one set of *F0* hypotheses.

3.1. Front end

3.1.1. Extracting hidden partials

Extracting hidden partials is essential to increase the accuracy of analyzing polyphonic signals with limited resolutions. As shown in the top plot of Figure 1, a peak of unsymmetric form might correspond to overlapped partials.

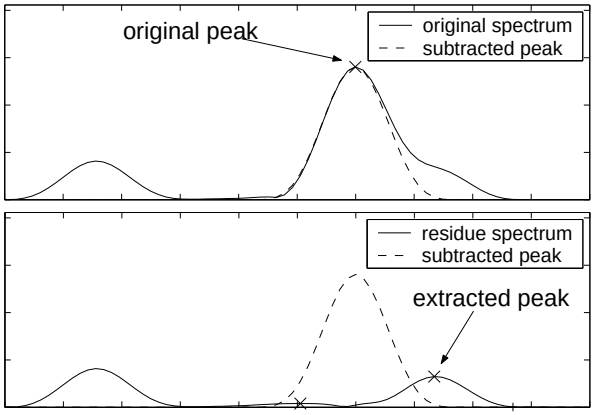

Figure 1: *Extracting the hidden partial*

To search for these hidden partials, we use a simple symmetry test for the shapes of the observed peaks. After selecting the peaks of relatively unsymmetric form and estimating their frequencies and frequency slopes [13], we subtract them one by one using the least square error criterion to extract the hidden peaks as the example shown in the bottom plot of Figure 1. To prevent the addition of simple residual energy as a new sinusoid, a resolved peak is kept as a successfully extracted partial only if it is not weaker than the original peak by 40 dB. Furthermore it should be located further than half the mainlobe width away from the original peak.

3.1.2. Generating the candidate list

To generate a *F0* candidate list, we use a harmonic matching technique because harmonicity is the primary concern in *F0* estimation. The harmonic matching technique matches the regular spacing between adjacent partials to determine a coherent *F0* and has been widely used for *F0* estimation in the spectral domain [9].

Given a *F0*, we construct a vector d_{F0} evaluating the degree of deviation from a harmonic model to the observed

peaks. A tolerance interval around each harmonic is used to measure the goodness of harmonic match. For the i-th observed peak matching the h-th harmonic, the degree of deviation is formulated as

$$d_{F0}(i) = \frac{|f_{peak}(i) - f_{model}(h)|}{\alpha \cdot f_{model}(h)} \quad (2)$$

where $f_{peak}(i)$ is the frequency of the i-th observed peak, $f_{model}(i)$ is the frequency of the hth harmonic of the model, and α determines the tolerance interval[1] allowing certain inharmonicity. If an observed peak situates outside the corresponding tolerance interval, it is regarded as unmatched and $d_{F0}(i)$ is set to 1. Since inharmonicity exists in most of the string instruments, it is necessary to dynamically adapt the frequencies of model harmonics according to the matched peaks. Thus, $f_{model}(h)$ is calculated by means of adding *F0* to the previously matched peak frequency. If not a single peak is matched for the previous partial, $f_{model}(h-1)+F0$ is used for the current match. The technique of selecting one single matched peak is described later.

Three vectors are chosen to weight d_{F0}: (i) the complex correlation between each observed peak and an ideal peak given by the mainlobe of the Fourier transform of the analysis window, (ii) the linear amplitudes of the observed peaks, and (iii) an attenuation vector favoring the first several partials and attenuating higher ones in proportion to h, as indicated in the top plot of Figure 2. According to empirical studies, the third partial is a good starting point for attenuation.

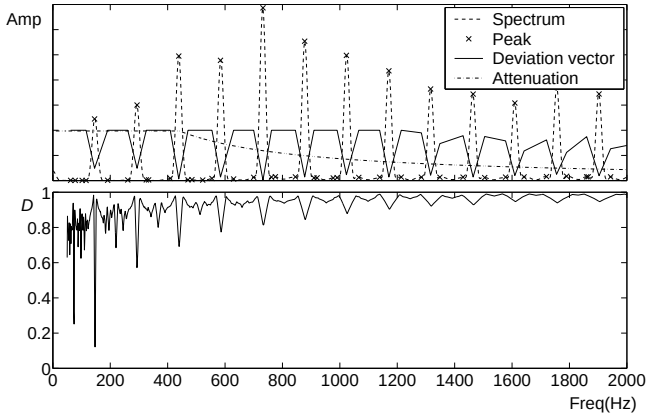

Figure 2: *Harmonic matching: a tenor trombone note at 137Hz*

The complex correlation favors peaks of better sinusoidality (shape and phase). The linear peak amplitude adjusts relative significance by considering peaks of larger energy more important. Spurious peaks are selected using the method in [12] and the corresponding weightings are set to zero, which functions as noise suppression. The third weighting vector attenuates less reliable matches for higher partials because they tend to be inharmonic and less stationary. Besides, the gradual decay nature of higher partials reduces the reliability in the presence of stronger partials from other sources. Then the weighted deviation vector is summed and normalized between 0 and 1. The resulting indicator for the harmonic match is denoted as D. An example is shown in the bottom plot of Figure 2, the values of D for *F0* hypotheses ranging from 50Hz to 2000Hz are plotted. A lower value means a better match and thus higher harmonicity. The harmonic match indicator is applied to polyphonic spectra and the *F0* hypotheses corresponding to local minima of D are selected as the candidates for joint evaluation.

Assume there are P *F0* candidates and M target *F0*s to be estimated from the observed spectrum, which results in the need to evaluate C_M^P hypothetic combinations of *F0* hypotheses.

3.1.3. Generating Hypothetic Partial Sequences

Constructing *HPS*s of *F0* hypotheses in the candidate list is realized by the partial selection technique. A *F0* hypothesis may have different *HPS*s in different hypothetic sets. Both Parsons [11] and Duifhuis [7] have proposed selecting the nearest peak around a harmonic. However, this technique might fail if a partial is surrounded by spurious peaks and partials of other sources. Therefore, we try to increase the robustness by means of utilizing **Principle 2** and the knowledge of spectral locations where partial overlaps may occur according to the set of *F0* hypotheses under investigation. The goal is to ensure that *HPS*s contain credible information for further evaluation.

To construct a *HPS* we start with the first partial by simply assigning it to the closest peak observed. For the following partials we consider two candidate peaks: the closest one and the one of which the mainlobe contains the corresponding harmonic position. Compared to the formerly selected partials, the peak candidate forming a smoother envelope search path is sequentially allocated to the *HPS*.

The case of overlapped partials requires special consideration. The treatment for this case is based on the idea that an overlapped partial still carries important information for at least the *HPS* that locally has the strongest energy. Therefore, the algorithm aims to select this *HPS* to assign the partial. Partials having potential collision are determined from the hypothetic set of *F0*s and the local energy strength of the envelope is obtained by means of interpolation of the neighboring partials that are not collided. For the rest of the *HPS*s the overlapped partial is labeled as existing but without a specified partial amplitude. The score criteria explained in the following are designed to gracefully deal with this kind of incomplete *HPS*s.

3.2. The score function

Having constructed the most reasonable *HPS*s for each set of *F0* hypotheses, we design a score function to rank these hypothetic sets. The score function formulates the three principles into four criteria: harmonicity *HAR*, mean bandwidth

[1] Two times the denominator in eq.(2)

MBW and duration *DUR* of *HPS*s, and the standard deviation of mean time *DEV*.

Criterion 1 *HAR* is an indication of harmonicity and totally "explained" energy. It is formulated as

$$HAR = \sum_{i=1}^{I} \frac{Corr(i) \cdot Spec(i) \cdot d_M(i)}{\sum_i [Corr(i) \cdot Spec(i)]} \tag{3}$$

where I is the number of peaks, $Corr$ is the complex correlation weighting vector, $Spec$ is the linear peak amplitude vector and $d_M(i)$ is obtained by combining $\{d_{F0_m}(i)\}_{m=1}^{M}$ at the i-th peak in the following way:

$$d_M(i) = min\big(\{d_{F0_m}(i)\}_{m=1}^{M}\big) \tag{4}$$

That is, each observed peak is matched with the closest partial among those of $\{HPS_{F0_m}\}_{m=1}^{M}$ and thus each combination under evaluation could perform its optimal match. The function of *HAR* is to prevent superharmonic errors.

Criterion 2 To evaluate the smoothness of a *HPS*, we use the mean bandwidth as a criterion. Each *HPS* is assembled with its "mirror sequence" to construct a new sequence S_{F0_m} for further evaluation. An example of S_{F0_m} is shown in the middle plot of Figure 3.

Applying K-point fast Fourier transform to S_{F0_m} to obtain the linear spectrum X_{F0_m}, we can calculate the mean bandwidth MBW_{F0_m} as

$$MBW_{F0_m} = \sqrt{2 \cdot \frac{\sum_{k=1}^{K/2} k[X_{F0_m}(k)]^2}{\sum_{k=1}^{K/2} [X_{F0_m}(k)]^2}} \tag{5}$$

This indicates the degree of energy concentration in low frequency region and thus S_{F0_m} with less variation results in a smaller value of MBW_{F0_m}.

The function of *MBW* is to discriminate correct *F0*s from subharmonics. As the example shown in Figure 3 the spectral envelopes of a clarinet note. Although the nature of the clarinet does not form a smooth spectral envelope due to the absence of even partials, the *HPS* of its subharmonic *F0/2* contains even more variations.

Criterion 3 For a quasiharmonic sound, the spectral centroid tends to lie around lower partials because higher partials often decay gradually. From this general principle related to **Principle 2**, we could similarly evaluate the energy spread of a *HPS*, that is, the duration DUR_{F0_m} of HPS_{F0_m}. Instead of removing the non-reliable components from HPS_{F0_m}, we simply set them to zero to maintain correct positioning of all partials. Then the duration of HPS_{F0_m} could be calculated as

$$DUR_{F0_m} = \sqrt{2 \cdot \frac{\sum_{n=1}^{N_m} n[HSP_{F0_m}(n)]^2}{L \cdot \sum_{n=1}^{N_m} [HSP_{F0_m}(n)]^2}} \tag{6}$$

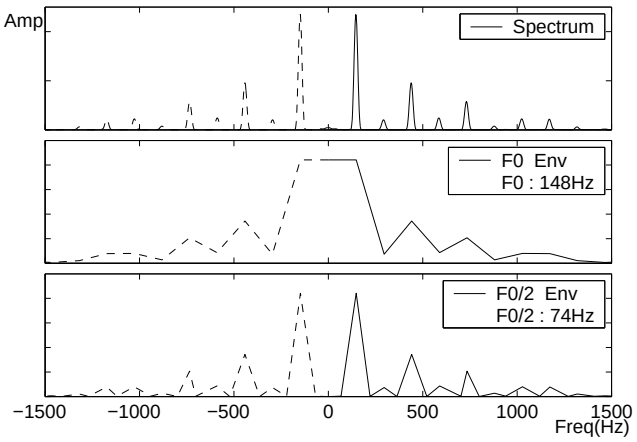

Figure 3: *Spectral envelope comparison between F0 and F0/2 of a clarinet note at 148 Hz*

where N_m is the length of HSP_{F0_m}. L is a normalization factor determined by $\lfloor F_{90}/F0_{min} \rfloor$, where F_{90} stands for the frequency limit containing 90% of spectral energy in the analyzing frequency range and $F0_{min}$ is the minimal hypothetic *F0* in search. Since spectral envelopes of natural sounds are not always smooth, this criterion functions as the further test of physical consistency of **Principle 2** and acts as a penalty function for subharmonics which "explain" more than one source in the observed spectrum.

Criterion 4 To evaluate the synchronicity of the temporal evolution of the hypothetic sinusoidal components in a *HPS*, we rely on the estimation of the mean time for individual spectral peaks. Mean time is an indication of the center of gravity of signal energy[2] and the mean time of a spectral peak can be used to characterize the amplitude evolution of the related signal[14]. For a coherent *HPS* we expect synchronous evolution resulting in a small variance of mean time concerning a collection of peaks.

The mean time of a hypothetic source, denoted as T_{F0_m}, is calculated as the power spectrum weighted sum of the mean time of the hypothetic partials. The variance of mean time of the partials in HPS_{F0_m} is then formulated as

$$VAR_{F0_m} = \sum_{i=1}^{I} \{[\bar{t}_i - T_{F0_m}]^2 \cdot w_{F0_m}(i)\} \tag{7}$$

where $\bar{t}_i$ denotes the mean time of the i-th observed peak and the weighting vector $\{w_{F0_m}(i)\}_{i=1}^{I}$ is constructed from HPS_{F0_m} by setting zeros for the following components: (i) non-reliable partials due to overlaps and (ii) close partials of which spectral phases are probably disturbed. Lastly, $\{w_{F0_m}(i)\}_{i=1}^{I}$ is compressed by an exponential factor to reduce the dynamic range such that the significance of spurious peaks is raised. This makes use of the spurious peaks to penalize more a *HPS* containing more spurious peaks. VAR_{F0_m} is then square-rooted and then normalized

by half of the window size to define DEV_{F0_m}.

Then *MBW* of a set of *F0* hypotheses is defined as the weighted sum of $\{MBW_{F0_m}\}_{m=1}^{M}$:

$$MBW = \frac{\sum_{m=1}^{M}[\sum_{n=1}^{N_m} HPS_{F0_m}(n)] \cdot MBW_{F0_m}}{\sum_{m=1}^{M}\sum_{n=1}^{N_m} HPS_{F0_m}(n)} \quad (8)$$

DUR and *DEV* are thus equivalently defined.

Score function The final score function is defined as

$$D_{C_M^P} = \frac{1}{\sum_{j=1}^{4} p_j}\{p_1 \cdot HAR + p_2 \cdot MBW + p_3 \cdot DUR + p_4 \cdot DEV\} \quad (9)$$

where $\{p_j\}_{j=1}^{4}$ are the weighting parameters for the four criteria. These criteria are designed in a way that a smaller weighted sum stands for higher score. The best of total C_M^P combinations has the highest score. Notice that *HAR* favors lower hypothetic *F0*s while *MBW*, *DUR* and *DEV* favor higher ones. Therefore, the criteria perform in a complementary way and the weighting parameters should be optimized to balance the relative contribution of each criterion such that the score function generally supports correct *F0*s the best.

4. Experimental results

To evaluate the proposed *F0* estimation method, we perform a frame-based test using mixtures of musical samples. Non-transient parts of musical samples are pre-selected and then mixed with equal mean-square energy. Estimation of a polyphonic sample is performed within a single frame. The number of *F0*s is given in advance for the *F0* estimation system to find the most probable set of *F0* hypotheses.

4.1. Parameter optimization

The parameters to be optimized are the weighting parameters $\{p_j\}_{j=1}^{4}$ in the score function and α for determining the tolerance interval in eq(2). 300 polyphonic samples containing 100 samples for each voice mixture are generated by randomly mixing musical instrument samples from the University of Iowa[2]. Then the parameters are optimized utilizing the evolutionary algorithm [15] and the set of parameters of the best performance ($\{p_j\}_{j=1}^{4} = \{20, 11, 11, 11\}, \alpha = 0.035$) is used for the final evaluation on a large database.

4.2. Evaluation setups and results

Specifications for this evaluation are described below:

- Three databases: two-voice, three-voice and four-voice mixtures, labeled as TWO, THREE and FOUR respectively, are generated using McGill University Master Samples[3]. In combining M-voice polyphonic samples, M out of twelve note names are preliminarily assigned and then samples ranging from 65Hz to 1980Hz are randomly mixed. Around 1500 samples for each database are generated in a way that each combination of note names are of equal proportion. Musical instrument samples not fitting the quasiharmonic model are excluded. To facilitate comparison, the database is published on the author's web page [4].

- The search range for *F0* is set from 50Hz to 2000Hz and the observed spectrum is analyzed up to 5000Hz. A Blackman window is used for analysis.

- *F0* reference values are created from single *F0* estimation of monophonic samples before mixing. A correct estimate should not deviate from the corresponding reference value by more than 3%. The error rates are computed as the number of wrong estimates divided by the total number of target *F0*s.

The results of evaluation using two analysis window sizes, 186ms and 93ms, are shown in Table 1 and Table 2, respectively. Since musical samples mixed randomly surely contain harmonically related notes, we present the error rates for two groups of samples: one group of mixtures containing harmonically related notes, labeled as "harmonical", and another group "non-harmonical". The overall error rates are shown in the "total" columns. The percentages of samples in the group "harmonical" are 22.43%, 32.78% and 49.46% for the three databases TWO, THREE and FOUR.

polyphony	non-harmonical	harmonical	total
TWO	0.58%	7.28 %	2.09%
THREE	1.48%	5.16 %	2.68%
FOUR	2.46%	6.57 %	4.50%

Table 1: *F0* estimation results using a 186 ms window

polyphony	non-harmonical	harmonical	total
TWO	1.61%	7.59%	2.96%
THREE	3.27%	7.61%	4.69%
FOUR	5.68%	11.78%	8.70%

Table 2: *F0* estimation results using a 93 ms window

The errors in the group non-harmonical are quite small which proves the satisfying performance of the proposed method. The overall errors are slightly better than the ones reported by Klapuri [10], however, this comparison is not conclusive due to the fact that the testing set comprises different samples and that in [10] a larger set of samples from four different databases has been used.

[2] http://theremin.music.uiowa.edu/MIS.html

[3] http://www.music.mcgill.ca/resources/mums/html/

[4] http://www.ircam.fr/anasyn/cyeh/database.html

5. Discussions

The score function sometimes fails to correctly resolve the ambiguity concerning target *F0*s and their subharmonics/superharmonics especially *F0/2* and *2F0*. This failure scenario accounts for a great proportion of the estimation errors. Polyphonic samples mixed with musical instrument samples of rich resonances often result in this kind of wrong estimates. Taking the string instruments for example, several predominant resonances occur with the excitation [8]. If strong resonances exist in the frequency range below the fundamental, the correct *F0*s might lose too much score to subharmonics by the amount of explained energy(*HAR*). If strong resonances boost certain partials too much, correct *F0*s might lose too much score to superharmonics by the spectral smoothness(*MBW*). Dealing with resonance peaks is a key to improving robustness.

With the increase of polyphony, the performance suffers from the reduction of the window size. Therefore, further improvements of the techniques for treating overlapped partials are necessary.

The way of constructing polyphonic databases for evaluation should be carefully examined. With the increase of polyphony, the number of possible combinations among different notes and different instruments increases dramatically. A limited number of samples mixed in a random manner could not ensure a general representation of the large sample space. Besides, the number of harmonically related notes increases in higher polyphonic random mixtures and thus effective approaches to estimate *F0*s of exact multiple relations become more important.

6. Conclusions

As discussed in [9], a successful *F0* estimation algorithm should make use of the principles which have complementary performances and combine them in an appropriate way. Following this proposal we design a new score function for joint evaluating the plausibility of multiple *F0* hypotheses. Evaluation over a large polyphonic database has shown encouraging results. However, there are still issues to be addressed. Despite the extraction of hidden peaks having been implemented to gain spectral resolution, the proposed method still suffers performance degradation due to an insufficient window size, especially for higher polyphony cases. Therefore, we envisage that further improvements on the inadequate treatment for overlapped partials will lead to higher robustness.

7. References

[1] Albert S. Bregman. *Auditory Scene Analysis*. The MIT Press, Cambridge, Massachusetts, 1990.

[2] Loen Cohen. *Time-Frequency Analysis*. Prentice Hall, 1995.

[3] M. Davy and S. Godsill. Bayesian harmonic models for musical signal analysis. In *Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting*, Valencia, Spain, 2003.

[4] Alain de Cheveigné. Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing. *Journal of Acoustical Society of America*, 93(6):3271–3290, 1993.

[5] Alain de Cheveigné and Hideki Kawahara. Multiple pitch estimation and pitch perception model. *Speech Communication 27*, pages 175–185, 1999.

[6] Boris Doval and Xavier Rodet. Estimation of fundamental frequency of musical sound signals. In *Proc. IEEE-ICASSP 91*, pages 3657–3660, Toronto, 1991.

[7] H. Duifhuis and L.F. Willems. Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception. *Journal of Acoustical Society of America*, 71(6):1568–1580, 1982.

[8] N. F. Fletcher and T. D. Rossing. *The Physics of Musical Instruments*. Springer-Verlag, New York, 2nd. edition, 1998.

[9] Wolfgang Hess. *Pitch Determination of Speech Signals*. Springer-Verlag, Berlin Heidelberg, 1983.

[10] Anssi Klapuri. *Signal processing methods for the automatic transcription of music, Ph.D dissertation*. Tampere University of Technology, 2004.

[11] Thomas W. Parsons. Separation of speech from interfering speech by means of harmonic selection. *Journal of Acoustical Society of America*, 60(4):911–918, 1976.

[12] A. Röbel and M. Zivanovic. A new approach to spectral peak classification. In *Proc. of the 12th European Signal Processing Conference (EUSIPCO)*, 2004. To appear.

[13] Axel Röbel. Estimating partial frequency and frequency slope using reassignment operators. In *Proc. of the International Computer Music Conference (ICMC'02)*, pages 122–125, Göteborg, 2002.

[14] Axel Röbel. Transient detection and preservation in the phase vocoder. In *Proc. Int. Computer Music Conference (ICMC'03)*, pages 247–250, Singapore, 2003.

[15] Hans-Paul Schwefel. *Evolution and Optimum Seeking*. Wiley & Sons, New York, 1995.

[16] M. Wu, D.L. Wang, and Brown G.J. A multipitch tracking algorithm for noisy speech. *IEEE Transactions on Speech and Audio Processing*, 11(3):229–241, 2003.