SSM – A Novel Method to Recognize the Fundamental Frequency in Voice Signals

György Várallyay Jr.

Budapest University of Technology and Economics,
Dept. of Control Engineering and Information Technology,
Magyar tudósok krt. 2. IB. 311, H-1117 Budapest, Hungary
varallyay@iit.bme.hu

Abstract. Nowadays the detection of the fundamental frequency (F_0) in voice signals can be evaluated by several algorithms. There are two main attributes of these algorithms: exactness and calculation time. A considerable part of the algorithms are based on the well-known Fast Fourier Transformation (FFT). The Smoothed Spectrum Method is an FFT based process, which was developed for the F_0 detection of recorded voice signals especially the infant cry. As it will be shown the SSM provides a better accuracy than regular FFT based algorithms or the Autocorrelation Function. In case of sound recordings in noisy environment the modified SSM is able to recognize significant noise components in the recorded signal. A further advantage of SSM is that additional information of the analyzed signal can be given to improve the performance of the method.

Keywords: Fundamental frequency detection, voice signals, noise detection.

1 Introduction

The fundamental frequency (F_0) is the lowest useful frequency component in the spectrum. The detection of the fundamental frequency has several applications, *e.g.* in mechanical engineering, in acoustical engineering, etc. In each of these applications various requirements are defined: robustness, calculation time, accuracy, etc [1-3]. In this study a novel method, named *Smooth Spectrum Method* (SSM) will be introduced, which was developed for the F_0 detection of *recorded voice signals* especially the infant cry. This method is based on the spectrum obtained by the well-known *Fast Fourier Transform* (FFT) [4].

Different types of voice signals result different spectrums. Figure 1 (a) shows the short-term spectrum of a sinusoid. An obvious way to estimate its fundamental frequency is to measure the position of the spectral peak. However this procedure fails for the spectrum in Fig. 1 (b) that contains multiple peaks. A simple modification is to accept only the largest peak, but this algorithm fails for the spectrum in Fig. 1 (c) for which the largest peak falls on a multiple of F_0. A simple extension is to select the peak of lowest frequency but this algorithm fails for the signal illustrated in Fig. 1 (d) for which the lowest peak falls on a higher harmonic. Another cue, spacing between partials indicates the correct F_0 for this signal, but not for the signal illustrated in

L. Rueda, D. Mery, and J. Kittler (Eds.): CIARP 2007, LNCS 4756, pp. 88–95, 2007.

Fig. 1 (e). In case of recorded acoustic signals narrow-band noises, and/or wide-band noises, and/or significant frequency components (f) might be added into the spectrum from background noises, the noise from the recording device, etc. It might result that some of the useful frequency components of the recorded signal under the noise level will disappear from the spectrum, while significant noise peaks can be treated as a useful frequency component.

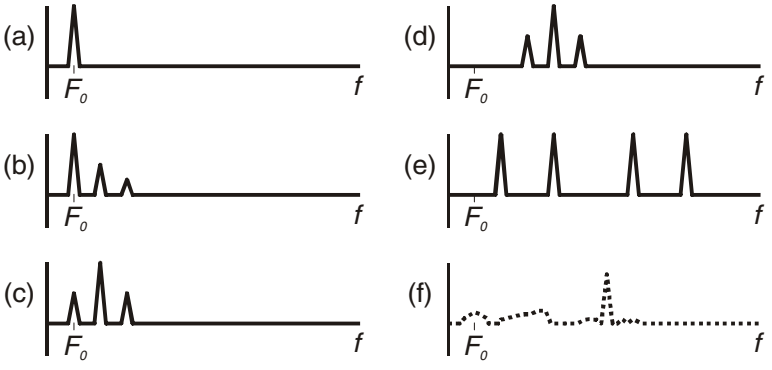

Fig. 1. Spectra of simple signals that illustrate basic spectral F_0 estimation schemes. The spectrum peak determines the F_0 of a pure tone (a) but a complex tone (b) has several such peaks. The largest peak determines the F_0 of the waveform in (b), but not (c). The lowest frequency peak determines the F_0 of the waveform of (c) but not (d). Interpartial spacing determines the F_0 of (d) but not (e). In case of recorded acoustic signals narrow-band noises, and/or wide-band noises, and/or significant frequency components might be added as it is shown in (f).

The inputs of the Smoothed Frequency Method are spectrums obtained from sound recordings of the infant cry. Cry is a multimodal, dynamic behavior; this is the first tool of communication and the sign of life at birth [5]. There are several purposes and ways to analyze the sound of crying: acoustic, physiological, psychological, phonetic, pediatric, etc [6-12]. The infant cry, on the analogy of voice signals, is mostly a harmonic signal, containing the fundamental frequency and its multiple integers, *i.e.* the subharmonics. The fundamental frequency of crying is typically between 200 and 800 Hz [13], while subharmonics can be found for 6000-8000 Hz. The amplitudes of these frequency components are different; there are significant components as well as missing ones, depending on the formant structure of the sound. [14]. In the acoustic analysis of the infant cry the maximum frequency of interest is often above 8000 Hz to be able to analyze special unvoiced sound phonemes.

As the Smoothed Spectrum Method is suitable not only for the F_0 detection of the infant cry but for the F_0 detection of other voice signals, in the following the inputs will be mostly harmonic signals (200 Hz $<F_0<$800 Hz), containing narrow-band and/or wide-band noises and/or significant noise components.

The sampling frequency applied was 44100 Hz; there was a window size of 2048 points. This window length (46.4340 ms) is typical in case of speech processing

or speech recognition, and provides a suitable resolution in the time domain. In the following these values will be used.

In regular FFT based algorithms the frequency resolution of the discrete spectrum is limited. In case of the window length above, the resolution of the spectrum is 21.5333 Hz. The frequency resolution of the FFT spectrums can be decreased theoretically *e.g.* by using a longer window, by zero padding, when possible and allowed. The Smoothed Spectrum Method is such a novel method for fundamental frequency detection, which provides a better accuracy than it could be reached theoretically by regular FFT based algorithms.

2 Methodology

The SSM can be divided into two consecutive parts. In the first part the input spectrum is smoothed and the significant peaks are detected. In the second part the most probable value of the fundamental frequency is calculated by statistical methods.

First, the input FFT spectrum is smoothed by a suitable (*e.g.* bell shape), symmetric kernel function. This smoothing can be realized by weighted addition in a predetermined bandwidth. The purpose of this step is to emphasize the significant peaks in the spectrum (*i.e.* the harmonic components) and to reject the wide-band noises in the spectrum. Figure 2 shows an original spectrum and its smoothed version.

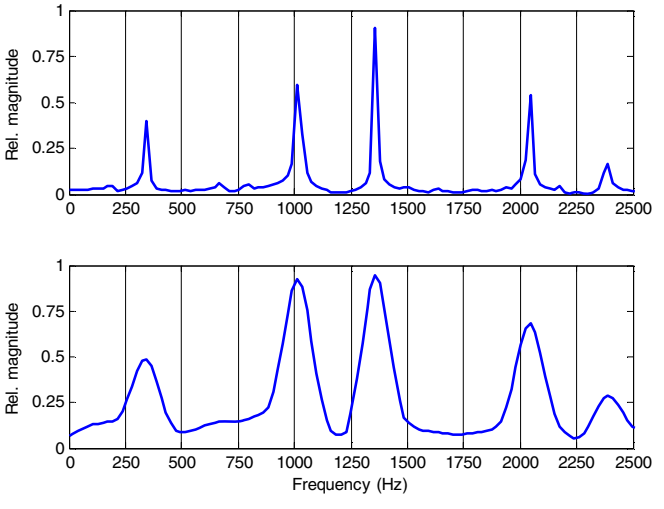

Fig. 2. An original spectrum (A) and its smoothed version (B). The significant peaks of the original spectrum are preserved, while the smaller (noise) components are rejected.

By the detection of the local maxima in the smoothed spectrum the significant peaks of the original spectrum are found. Note that the smoothing of the spectrum does not change the frequency position of the significant peaks.

The second part of the Smoothed Spectrum Method is based on these obtained peaks. As it was shown before there are several types of spectrums. In some cases the

fundamental frequency or higher frequency components are missing, or some significant noise components are present. First, let us see the cases without any noise components.

2.1 SSM Without Noise

In ideal case, the subharmonics (F_{n_ideal}) are whole multiples of the fundamental frequency (F_0).

$$F_{n\_ideal} = F_0 \cdot n \tag{1}$$

From this expression, if a subharmonic of F_0 is detected and its serial number is known, the fundamental frequency can be calculated by a simple division:

$$F_0 = F_{n\_ideal} / n \tag{2}$$

But the distance between two neighboring points in the discrete spectrum is

$$B = F_s / N \tag{3}$$

where F_s is the sampling frequency and N is the window length applied at the FFT. In the worst case a spectral component (F_x) of the signal is placed to $F_x \pm B/2$. In this case (1) is modified as

$$F_n = F_0 \cdot n \pm B/2 \tag{4}$$

From (2) and (4)

$$F_n / n = F_0 \pm \frac{B}{2n} = F_0 \pm \frac{F_s}{2nN} = F_0 \pm h = F_0\left(1 \pm h'_n\right) \tag{5}$$

Where h is the absolute error of the calculation of F_0, and h' is the relative error:

$$h'_n = \frac{F_s}{2nNF_0} \tag{6}$$

As the sampling frequency (F_s) and the window length (N) are given with the input spectrum, the relative error can be decreased by applying higher n value, *i.e.* determining higher harmonic component in the spectrum.

The significant peaks in the input spectrum are the local maxima in the smoothed spectrum. The positions of these peaks are close to the real subharmonics of F_0. If the frequency values of these peaks are divided by their serial numbers, the resulted ratios will be close to the real F_0 with the relative error from (6).

Example from Figure 2: the fundamental frequency is 340 Hz, the second and the fifth subharmonics are missing. As the frequency resolution of the spectrum (B) is 21.53 Hz, the detected peaks in the smoothed spectrum are at F_{det1}=335.47 Hz, F_{det2}=1027.94 Hz, F_{det3}=1363.41 Hz, F_{det4}=2034.35 Hz and F_{det5}=2369.81 Hz. After the division with their serial numbers the resulted ratios are: 335.47, 342.64, 340.85,

339.06 and 338.54 Hz. The differences between these results and the exact value of the fundamental frequency are within the theoretical error bands.

For these divisions the exact serial numbers of the detected peaks are needed. How could these serial numbers be obtained? By the combination of possible serial numbers the algorithm generates test sequences, and the ratios of the obtained peaks and all of these sequences are calculated. In case of the best sequence of serial numbers the standard deviation of the obtained ratios will be the least.

Continuing the previous example, five test sequences are given, and the ratios are calculated (see Table 1). As the standard deviation of these ratios has the smallest value at the second case, the best sequence of serial numbers is <1,3,4,6,7>.

Table 1. Examples of the sequences of serial numbers. As the standard deviation of the ratios is the smallest in the second line, the best combination is <1,3,4,6,7>.

Sequence of serial numbers	Ratios of the detected peaks and possible sequence of serial numbers (Hz)					Standard deviation (Hz)
	F_{det1}	F_{det2}	F_{det3}	F_{det4}	F_{det5}	
<1,2,3,4,5>	335.47	513.97	454.47	508.59	473.96	72.41
<1,3,4,6,7>	335.47	342.65	340.85	339.06	338.54	2.69
<1,2,4,6,8>	335.47	513.97	340.85	339.06	296.23	85.22
<2,3,4,5,6>	167.73	342.65	340.85	406.87	394.97	95.83
<2,3,4,6,7>	167.73	342.65	340.85	339.06	338.54	77.18

As the least value of the standard deviations is found, the fundamental frequency of the signal can be calculated by the division of the highest detected frequency component and its serial number. In this example the original F_0 at 340 Hz was positioned in the spectrum to 335.47 Hz, while a better estimation (338.54 Hz) was calculated with the SSM.

2.2 SSM with Narrow-Band and/or Wide-Band Noises

In case of wide-band noises, some of the useful peaks in the input spectrum might not be visible; but the Smoothed Spectrum Method can detect the fundamental frequency from the remaining significant peaks.

In case of narrow-band noises, or especially significant noise components, the smoothing of the spectrum might not reject these components. If such a noise component is expected in the signal, the algorithm of the SSM can be modified to recognize the extraordinary peak. In the modified algorithm divisions are evaluated not only with the detected peaks (including the noise peak), but with smaller groups of the detected peaks as well. There will be at least one group, where no noise peak is present.

An example for the recognition of a significant noise component: supposing that the fundamental frequency is 60 Hz, and in the signal there are significant peaks around 40, 60, 120, 180 and 300 Hz. In this example the significant noise component is at 40 Hz. The divisions are evaluated first with all the peaks, second with smaller

Table 2. An example for the recognition of a noise component. The harmonic signal has components at whole multiples of 60 Hz, while an extra peak at 40 Hz illustrates the significant noise component. When the noise component is skipped, the resulted fundamental frequency value differs from the others. That is the indicator of the noise component.

Detected peaks (Hz)	The best sequence of serial numbers after simple SSM	The calculated value of the fundamental frequency (Hz)
40, 60, 120, 180, 300	<2,3,6,9,15>	20
60, 120, 180, 300	<1,2,3,5>	60
40, 120, 180, 300	<2,6,9,15>	20
40, 60, 180, 300	<2,3,9,15>	20
40, 60, 120, 300	<2,3,6,15>	20
40, 60, 120, 180	<2,3,6,9>	20

groups of the peaks. See Table 2 how the modified SSM algorithm detects a noise component.

However the recognition of a noise component needs more SSM divisions, the modified algorithm is able to select the extraordinary peak from the significant peaks detected in the smoothed spectrum.

3 Comparison

To test the accuracy of the Smoothed Spectrum Method, harmonic signals were generated randomly and several fundamental frequency detection algorithms were applied and their results were compared.

There were constant values, as the sampling frequency (F_s=44100 Hz) and the length of the signal (N=2048 points), which resulted a B=21.53 Hz frequency resolution in the FFT spectrum. The fundamental frequency of the generated test signals were integers between 200 and 800 Hz. The amplitudes of the frequency components were randomly generated in each case; a maximum of 10 subharmonics were present.

In the comparison, besides the Smoothed Spectrum Method (SSM) the *Autocorrelation Function* (XCOR) and *Regular FFT* (RFFT) algorithms were applied. The Autocorrelation Function is a special type of the cross-correlation function, which is a typical method for F_0 detection *e.g.* in speech recognition [1, 4]. The Regular FFT algorithms collectively use the resolution of the FFT spectrum; such algorithms are the local maximum detection in the spectrum within a predetermined frequency interval, and the Harmonic Product Spectrum method [2].

For each test signal the value of the fundamental frequency was determined by all of the algorithms mentioned above, and the differences between the exact F_0 and the detected F_0 values were calculated. After this, the mean value of the absolute differences and the standard deviation of the differences were obtained. The following table (Table 3) shows the accuracy of the compared methods.

Table 3. Comparison between the accuracy of F_0 detection algorithms. IDEAL: in ideal case; SSM: by the Smoothed Spectrum Method; XCOR: by the Autocorrelation Function; and RFFT: by Regular FFT algorithms.

	Difference between the original and the detected values on the average (Hz)	Standard deviation of the difference between the original and the detected values (Hz)
IDEAL	0.0000	0.0000
SSM	0.6427	0.7617
XCOR	1.6717	2.2149
RFFT	5.3852	6.2251

As it is shown by Table 3 after the ideal case the Smoothed Spectrum Method provides the least detection error. The Autocorrelation Function also has a better error band than what a Regular FFT algorithm could reach.

Note that in case of regular FFT algorithms, a uniform distribution is expected for the detection difference between 0 and $B/2$=10.7666 Hz, which has an average value at 5.3833 Hz.

4 Discussion and Conclusion

A novel algorithm for fundamental frequency detection of the infant cry was introduced in this study. However the Smoothed Spectrum Method was developed to detect the fundamental frequency of crying signals, its conceptions are relevant for several voice signals, as speech, singing, and musical instruments.

In summary, the algorithm of the Smoothed Spectrum Method is the following:

- Smooth the input spectrum by a chosen kernel function in a chosen bandwidth;
- Detect the local maxima in the smoothed spectrum within a chosen frequency range;
- Generate possible sequences of serial numbers for the detected peaks;
- Divide the peaks with these sequences and calculate the standard deviation of the resulted ratios;
- Choose the smallest standard deviation to find the exact sequence of serial numbers and to calculate the fundamental frequency.

A further advantage of the SSM is that additional information of the analyzed signal can be given to improve the performance of the method. Users can give:

- The type of the kernel function and the bandwidth of smoothing;
- The interval of the local maximum detection in the smoothed spectrum;
- Rules and limitations for generating sequences of serial numbers for the division;
- Rules for detecting and skipping noise-peaks from the division.

For exactness the Smoothed Spectrum Method is a promising algorithm for fundamental frequency detection of voice signals. The modified algorithm of the SSM is able to recognize and eliminate the significant noise components in the signal. The

disadvantage of the SSM might be the calculation time, because it contains numerous divisions.

Acknowledgements. This study is a part of a biomedical project dealing with early diagnostics, started in 2001 in Hungary. Besides the author further members of the team are Professor Zoltán Benyó and Professor András Illényi from the Budapest University of Technology and Economics; Zsolt Farkas and Gábor Katona chief doctors from the Heim Pál Hospital for Sick Children, Budapest. In this study the author discussed his own research in this project. This project is supported by Hungarian National Office for Research and Technology (NKTH).

References

1. Deller, J.R., Proakis, J.G., Hansen, J.H.L.: Discrete-time processing of speech signals. MacMillan Publishing Co, New York (1993)
2. Parsa, V., Jamieson, D.G.: A Comparison of High Precision F0 Extraction Algorithms for Sustained Vowels. J. of Speech, Language, and Hearing Research 42, 112–126 (1999)
3. Cheveigné, A., Kawahara, H.: YIN, a fundamental frequency estimator for speech and music. J. Acoust. Soc. Am. 111, 1917–1930 (2002)
4. Randall, R.B.: Application of B&K equipment to frequency analysis. Brüel & Kjaer Techn. Library, Denmark (1977)
5. Barr, R.G., Hopkins, B., Green, J.A.: Crying as a sign, a symptom and a signal. MacKeith Press, London (2000)
6. Hirschberg, J., Szende, T.: Pathological cry, stridor and cough in infants. Akadémiai Kiadó, Budapest (1982)
7. Várallyay Jr., G., Benyó, Z., Illényi, A.: The development of the melody of the infant cry to detect disorders during infancy. In: Proc. IASTED International Conference on Biomedical Engineering (BioMED 2007), Innsbruck, Austria, February 14-16, pp. 186–191 (2007)
8. Möller, S., Schönweiler, R.: Analysis of infant cries for the early detection of hearing impairment. Speech Commun. 28, 175–193 (1999)
9. Cacace, A.T., Robb, M.P., Saxman, J.H., Risemberg, H., Koltai, P.: Acoustic features of normal-hearing pre-term infant cry. Int. J. Pediatr. Otorhinolaryngol. 33, 213–224 (1995)
10. Várallyay Jr., G.: Future Prospects of the Application of the Infant Cry in the Medicine. Per. Pol. Elec. Eng. 50, 47–62 (2006)
11. Fort, A., Manfredi, C.: Acoustic analysis of newborn infant cry signals. Med. Eng. Phys. 20, 432–442 (1998)
12. Michelsson, K., Michelsson, O.: Phonation in the newborn, infant cry. Int. J. Pediatr. Otorhinolaryngol. 301, S297–S301 (1999)
13. Várallyay Jr., G.: Infant cry analyzer system for hearing disorder detection. Trans. on Automatic Control and Computer Science 49, 57–60 (2004)
14. Wermke, K., Mende, W., Manfredi, C., Bruscaglioni, P.: Developmental aspects of infant's cry melody and formants. Med. Eng. Phys. 24, 501–514 (2002)