2006-01-01

# Music Transcription Within Irish Traditional Music

Mikel Gainza
*Technological University Dublin*

## Recommended Citation

Gainza, Mikel (2006) *Music transcription within Irish traditional music.* Doctoral Thesis, Technological University Dublin. doi:10.21427/D7T03S

# Music Transcription within Irish Traditional Music

Mikel Gainza

A thesis presented to the Dublin Institute of Technology,

Faculty of Engineering,

For the degree of

Doctor of Philosophy

2006

Research Supervisors: Dr. Robert Lawlor

Dr. Eugene Coyle

I certify that this thesis which I now submit for examination for the award of Doctor of Philosophy, is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

This thesis was prepared according to the regulations for postgraduate study by research of the Dublin Institute of Technology and has not been submitted in whole or in part for an award in any other Institute or University.

The Institute has permission to keep, to lend or to copy this thesis in whole or in part, on condition that any such use of the material of the thesis be duly acknowledged.

Mikel Gainza

Mikel Gainza

Date: June 2006

# Abstract

Transcribing Irish traditional music is an open-field of research. The oral transmission of the music between generations explains the lack of transcription until recent times. The music can be played solo, which permits the player to exploit the variety of ornamentation types, in unison, and also with the accompaniment of a harmonic instrument. Different signal processing applications for transcribing Irish traditional music are presented in this thesis, including onset, ornamentation and pitch detection.

An onset detection system which focuses on the characteristics of the tin whistle within Irish traditional music is first presented. The tin whistle is a good example of the features of Irish traditional music, and the detection of its onset encounters all the problems associated with onset detection identified in the literature review. An extension of this method is also implemented in an effort to detect the most common types of ornamentation, which has not been attempted to date.

Existing onset detectors utilise energy and/or phase information to detect onsets. A novel onset detector, which focuses on the harmonicity of the signal to detect the onsets by using comb filters, is presented. This method overcomes the difficulties encountered by existing onset detection approaches in respect of signal modulations and detection of slow onsets. Finally, a further comb filter based method is utilised to detect the triads played by a harmonic accompaniment.

A set of results is presented for the four methods, followed by a commentary and explanation of the novel contributions.

# Acknowledgements

I would first like to thank Dr. Robert Lawlor for supporting this research and for facilitating my incorporation into the DIT audio research group. I am deeply grateful to Dr. Eugene Coyle for his overwhelming support, advice and eagerness to help in any way he could. I would also like to thank the Faculty of Engineering and the Faculty of Applied Arts at Dublin Institute of Technology and the interdisciplinary research team of the Digital Tools for Music Education, DiTME, research project.

I am indebted to Dr. David Dorran, whose incredible help in numerous ways throughout the completion of the PhD has been of enormous assistance to me. This included memorable and rewarding audio DSP discussions, in addition to reading through many of my draft documents and assisting in editing my English. The feedback has been of great benefit and I feel that it has helped significantly improve my writing. I should not forget his skill at being able to withstand my compulsive morning chatter!

I wish to extend this gratitude (in no particular order) to Dan Barry, Charlie Cullen and Dr. Derry Fitzgerald; for being something more than faithful proof-readers, at the moments when I was most in need. I am also grateful to all the members of the DMC, whose good humor and collaborative behavior made working there a very agreeable experience.

I am also thankful to Bryan Duggan, who consented to being recorded playing the Irish tin whistle and for the great discussions in ornamentation techniques we shared. Special

thanks to Juan Pablo Bello, for making the code available that made possible the onset detection comparison tests.

Deepest thanks go to my family, whose love and emotional support guided me during all the ups and downs throughout this PhD. In particular to my mother, my life teacher, who after all that listening and advising, I believe knows more about my PhD than I do!

Finally, I would like to thank Carmen Martinez, whose love, passion and joy illuminates the cloudiest day. *Sin tu cariño, nunca hubiera sido capaz.*

# Table of Contents

# Table of Figures

# Table of Tables

# 1 Introduction

## 1.1 Irish Traditional Music

The definition of Irish traditional music is very broad, and descriptions of its musical characteristics are open to more than one interpretation. Its scope accommodates both dance and non-dance music, and both instrumental and singing music. Irish traditional music is played everywhere: pub sessions, street corners and even in large shows such as the "Riverdance". The music is normally played in unison [Prout '06], but in recent years it has also seen the addition of harmonic accompaniment, as well as the modernist fusion with other music styles such as jazz music [Vallely '99].

The "traditional" adjective is derived from its oral tradition, and has passed between generations by listening and imitation, explaining to a certain extent the reason why a large amount of traditional musicians do not read music. However, nowadays Irish traditional music is widely taught by formal courses and the use of music notation is seen as a great tool for teaching and learning. In addition, experienced players often utilise it as a reminder to aid the musical memory. The recent incorporation of music notation partly explains why the development of signal processing applications for traditional musicians, professional or beginners, has not yet been explored.

Standard music notation (staff notation) uses symbols to indicate the onset time, pitch and note duration. Many music transcription systems utilise onset detectors in order to segment the signal, after which the pitch of the notes that comprise each musical segment are then calculated [Godsmark '99, Kashino '95b]. The duration of the note can be obtained by calculating the difference between the offset and the onset time. Additional fields such as the tempo or the key signature can be calculated by utilising onset and pitch information respectively [Scheirer '98]. This shows that the efficiency of onset detection algorithms is crucial in music transcription systems.

Ornamentation plays a very important role in Irish traditional music [O'Canainn '93]. However, it is interpreted differently in Irish traditional music than in classical music. In classical music, the expression is achieved by adding notes to the melody. By contrast, with the exception of the slide effects, Irish traditional music ornamentation is played on the beat, and alters the onset of the notes in a manner in which only one note will be heard (as opposed to two notes as in classical music) [Larsen '03].

Irish traditional music has slowly evolved over several hundred years, and instruments such as the banjo were not accommodated until the 1930s. Other instruments such as the guitar or bouzouki were not introduced till the 1960s (inspired by the American folk revival) and are yet to be fully accepted [Carson '99, Vallely '99]. However, the presence of the fiddle, the tin whistle and simple system flute, as well as the free reed instruments concertina and button accordion dominate in the majority of the tunes [Carolan '06]. All of these instruments have something in common; their onsets have a slow profile which takes some time to reach the maximum amplitude value, as opposed to sharp onsets typically found with percussive instruments such as the piano.

## 1.2 Aims and overview of the Thesis

The principal aim of this thesis is to develop different signal processing algorithms for the purpose of transcribing Irish traditional music. This includes onset, pitch and ornamentation detection. As mentioned in the previous section, onset detection systems are integrated into the majority of music transcription systems. Thus, the development of a robust system capable of detecting onsets within Irish traditional music represents a major part of the presented thesis.

In order to achieve this objective, the different approaches that perform onset detection are first reviewed. Existing onset detection approaches generally perform successfully on detecting sharp onsets. However, their performance considerably degrades if the onset has a slow profile, or when amplitude and frequency modulations are present in the signal.

In addition, a literature review of pitch detection approaches has also been undertaken. A pitch detection model that deals with the singularities of Irish traditional music such as ornamentation techniques has not yet been implemented. Of the existing techniques, knowledge based representations integrate musical information into signal processing algorithms. These systems generally utilise an onset detector to segment the signal prior to the pitch analysis, which inter-relates the onset and pitch detection problems.

The implementation of a robust system capable of detecting slow onsets is still an open issue. In order to deal with the problem, an energy based method which focuses on the characteristics of a given instrument within Irish traditional music is presented [Gainza '04c]. This method was customised to the characteristics of the tin whistle, which is according to [Vallely '99], "without doubt the most popular instrument in traditional music today". This is a good example of an instrument with a slow onset, and introduces both frequency and amplitude modulations, which cause difficulties to detect using existing onset detection techniques.

Onset detection systems produce a detection function, from which the onset candidates are picked by using a threshold [Bilmes '93]. In this thesis, three different novel thresholding methods are considered. The first method is based on the standard deviation method [Pal '05], the second sets the thresholds according to the expected blowing pressure that a tin whistle produces per note. Finally, a third thresholding method combines the first and second methods.

Based on the latter onset detection method, a novel algorithm that detects single-note ornamentation such as cuts and strikes has been developed [Gainza '04a]. In addition, multi-note ornamentation such as rolls and cranns are also identified [Duggan '06b] which completes the ornamentation transcription system.

The onset detector based on the tin whistle is also capable of dealing with notes played legato. In addition, by adequately thresholding the onset detection functions, the problems due to amplitude modulated signals are significantly reduced. However, the system is prone to errors caused by frequency modulations and strong amplitude modulations. In order to reduce the effect of the signal modulations, a new onset detector which utilises FIR Comb Filters is presented. This novel approach takes advantage of the harmonic shape of the comb filter spectrum in order to combine energy and harmonicity signal information. The results notably improve the accuracy over existing onset detectors when detecting slow onsets [Gainza '05b].

As outlined in the introduction, Irish traditional music has been historically played in unison. In this case, a monophonic pitch detector should be capable of detecting the note that has been simultaneously played by the performers, and polyphonic pitch detection will not be required. However, harmonic accompaniment has been recently added to Irish traditional music, and is generally performed by a stringed instrument such as a guitar or bouzouki. The effectiveness of comb filter techniques in detecting slow onsets by tracking harmonicity signal changes is shown in [Gainza '05b]. This leads to the use once more of comb filter techniques for estimating the pitch[1] of the notes that comprise the harmonic polyphonic mixture [Gainza '05a].

### 1.2.1 Summary of contributions

The attempt to transcribe Irish traditional music represents a novel contribution in the field, since transcribing this type of music has never been attempted previously. In

[1] The terms pitch and fundamental frequency *(f0)*, have historically been utlised as synonyms. However, there is a conceptual difference between the two terms, where pitch refers to the perceptual attributes of the fundamental frequency [Klapuri '98]. In order to be cohesive with the terminology utilised by the existing research in the area, the term pitch detection has been utilised throughout this thesis to describe the detection of the fundamental frequency. In addition, pitch detection also describes the detection of the inverse of the fundamental frequency, which is denoted as "pitch period" [Moorer '74].

addition, the main specific contributions to knowledge included in the presented thesis are listed as follows:

**Contribution 1** The novel development of a slow onset detector customised to the characteristics of the tin whistle. The model uses a multi-band configuration adapted to the notes and modes played by the tin whistle. The results show that the method improves upon existing onset detectors [Gainza '04c].

Three different novel thresholding methods have been implemented, from which two set the thresholds automatically. The two methods perform successfully in the onset detection system, improving upon existing automatic thresholding methods. The development of these thresholding methods can be interpreted as a sub-contribution within Contribution 1.

**Contribution 2** A novel ornamentation detector based on the system related to Contribution 1 has been implemented. The method transcribes the most widely played single and multi-note ornamentation types such as cuts, strikes, rolls and cranns [Gainza '04a].

**Contribution 3** The development of a novel onset detector which extracts the signal harmonicity structure by the use of comb filters. The system improves the accuracy upon existing methods on detecting slow onsets, and on dealing with frequency and amplitude modulations. In addition, the system provides a more accurate onset time [Gainza '05b].

**Contribution 4** The harmony provided by a musical accompaniment is captured by a multi-pitch estimator based on comb filters, which is utilised to detect the triads played by the accompaniment instrument. The system improves upon existing comb filter based multi-pitch detectors [Gainza '05a].

## 1.3 Contents of the Thesis

The research undertaken in this thesis is contained within the following chapters:

**Chapter 2: Irish Traditional Music** - this chapter documents the general aspects that describe Irish traditional music. This covers the main instruments played, the structure of the music and ornamentation theory. Due to the prevalence of the Irish tin whistle within Irish traditional music, its musical characteristics are also described. This knowledge is used in Chapters 6 and 7 to develop an onset detector and an ornamentation detector respectively.

**Chapter 3: Comb filtering** - this chapter provides a brief description of comb filter techniques. These methods have been utilised in this thesis to implement different musical applications within an Irish traditional music context, such as onset detection (Chapter 8) and pitch detection (Chapter 9).

**Chapter 4: Onset Detection** - this chapter reviews the different onset detection approaches, and discusses the advantages and disadvantages of the existing methods. The conclusions documented in this chapter leads to the development of the onset detection methods presented in Chapters 6 and 8.

**Chapter 5: Pitch detection** - this chapter reviews the different pitch detection approaches. A discussion of the methods is also given, investigating their use in an Irish traditional music context. The chapter serves as an introduction to the ornamentation and pitch detection methods develop in Chapters 7 and 9 respectively.

**Chapter 6: Onset Detection System applied to the Tin Whistle (ODTW)** - this chapter presents a novel onset detector customised according to the characteristics of the Irish tin whistle described in Chapter 2. The system is compared against existing onset detection methods.

**Chapter 7: Ornamentation transcription** – this chapter presents a novel ornamentation detector based on the theory introduced in Chapters 2 and the onset detector presented in Chapter 6.

**Chapter 8: Onset Detection system based on Comb Filters (ODCF)** - this chapter presents a novel onset detector based on FIR comb filters which focuses on the harmonicity of the signal. The system can be utilised to detect onsets of any slow onset instrument. The system is compared against existing onset detection methods and the ODTW.

**Chapter 9: Multi-pitch Estimation Using Comb Filters (MPECF)** – this chapter presents a novel multi-pitch detector based on comb filters. The system is compared against existing comb filter based pitch detection methods.

**Chapter 10: Summary and future work** – this chapter documents the main conclusions, and discusses further work.

# 2 Irish Traditional Music

As previously mentioned in the introduction chapter, Irish traditional music contains various definitions, musical contexts and forms. This chapter aims to introduce the general aspects that describe Irish traditional music. The main instruments that are part of Irish traditional music are documented in Section 2.1. Due the high relevance of the tin whistle within Irish traditional music, Section 2.3 is devoted to its musical characteristics. In this thesis, applications for detecting the onsets and transcribing the ornamentation played by the tin whistle are presented in Chapters 6 and 7, respectively. Section 2.2 focuses on the structure of the music, covering the modal nature of the music, its different forms and introducing the ornamentation. Finally, a discussion and some conclusions are given in Sections 2.4 and 2.5 respectively.

## 2.1 Instruments

As mentioned in the introduction, there are numerous instruments that currently play Irish traditional music. However, the bulk of the music is played by the fiddle, the uilleann pipe, the tin whistle and simple system flute, as well as the free reed instruments concertina and button accordion [Carolan '06].

### 2.1.1 The Fiddle

Even though the fiddle is exactly the same instrument as the violin, the style of playing differs considerably in traditional and classical music. As an example, the use of vibrato is practically non existent in traditional dance music, as opposed to classical music where it is an integral part of the style of the player [Carson '99, Vallely '99]. In Figure 2-1, a picture of a fiddle player performing along with a flute player is shown.

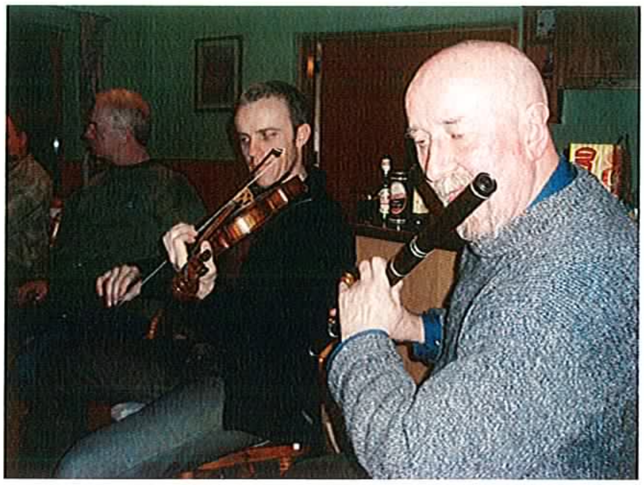

**Figure 2-1: Fiddle and flute players in a pub session**

### 2.1.2 The uilleann (elbow) pipe

The *uilleann* (elbow) *pipe* refers to the bellow-blown bagpipe, which supplies a continuous flow of air to the instrument. The melody line is supplied by a chanter which is usually tuned in D. In addition, three *drones* provide a constant accompaniment to the lowest note of the *chanter*, which are tuned in unison, one octave below and two octaves below respectively. Finally, keyed melody pipes (regulators) are capable of providing occasional harmony to the drones and chanter [Carson '99, Vallely '99]. It is paradoxical that one of the oldest instruments in a historically melodic music, has this significant potential to provide harmony accompaniment. It is documented that the regulators were added to satisfy ears of the nineteenth century musicians [O'Canainn '93]. However, pipers rarely exploit the harmonic possibilities of the instrument [Carson '99]. The uilleann pipe has been a very important instrument in the development of the styles of melodic instruments within the Irish traditional music, specially concerning the ornamentation techniques [Larsen '03].

### 2.1.3 The Simple-System Flute

The simple-system flute, also called Irish flute, is a mouth blown instrument with six holes that are exclusively covered by the fingers (as opposed to by a key mechanism) [Larsen '03]. As an example, Figure 2-1 displays a picture of a simple-system flute player performing in a traditional music session. The simple-system flute can also have additional holes covered by keys to extend the possibilities of the instrument. However, these keys are not needed to play the vast majority of Irish traditional music tunes.

### 2.1.4 Free Reed Instruments

Melody free reed instruments such as the *melodeon, button accordion* and the *anglo concertina* are also widely utilised in Irish traditional music. In these instruments, the air stream is generated by the action of blowing a bellow using the hands, which go across a set of paired metal reeds causing them to vibrate. The three instruments are single-action instruments, which have keys that can produce two notes depending if the player presses or draws the bellow [Vallely '99]. The melodeon has a set of ten keys, which produces twenty notes of the diatonic scale. The instrument was replaced by the button accordion, which includes a row of keys to produce a full chromatic scale. Since traditional music is essentially diatonic, the second row is reserved to produce ornamentations. Finally, the anglo concertina is a small accordion with hexagonal shape, having five keys at each side [Vallely '99].

### 2.1.5 Other instruments

The *tin whistle* is a six holes fipple flute from the family of the recorder. [Vallely '99] states that the Irish tin whistle "is without doubt the most popular instrument in traditional music today". In Section 2.3, a more detailed description of the Irish tin whistle is given.

Other instruments include the *harp*, which is the national symbol of Ireland. However, the decline that the instrument suffered in the seventeenth and eighteenth century fractured the oral transmission between generations [Vallely '99]. As a result, today's harpers are seen as innovators, and the harp is not fully integrated in Irish traditional music [Vallely '99].

Percussion in Irish traditional music plays a minor role [Carolan '06]. When utilised, the bodhrán would be a common choice. This instrument is a one single side frame drum made with goat skin, and it is played with the hand or a stick. An example of a bodhrán participating in a traditional session is illustrated in Figure 2-2.

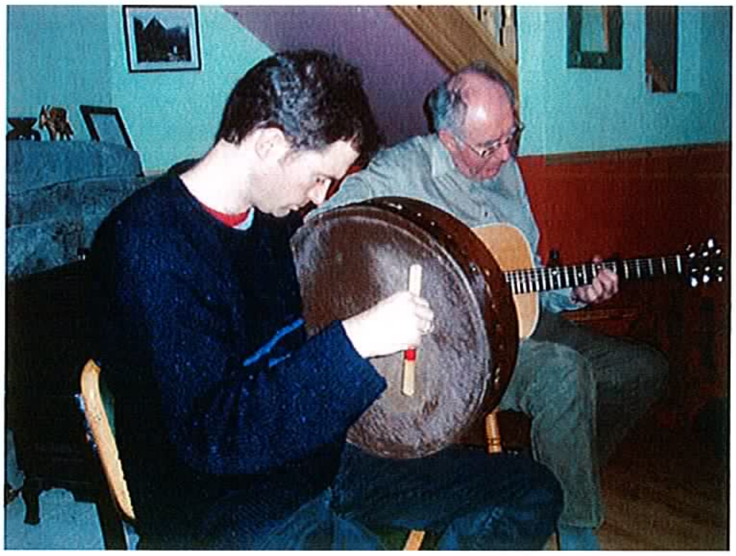

**Figure 2-2: Bodhrán and guitar accompaniment playing together**

When utilised, accompaniment would generally be of a simple kind. Dominant harmonic instruments are the *guitar* and the *bouzouki*, which were not introduced until the 1960s inspired by the American folk revival. In Figure 2-2, a guitar providing accompaniment is shown in the background. However, harmonic instruments are yet to be fully accepted in all traditional musician circles [Carson '99, Vallely '99]. During the present time, there is also a tendency of using a piano to provide harmony in contemporary commercial recordings. However, due to the bulky, heavy and expensive nature of the instrument, it is not used in informal non-commercial pub sessions.

## 2.2 Structure of the music

### 2.2.1 Modes

Irish traditional music has a modal nature, which means that the tones which compose the scales are based on the seven modes developed in the middle ages. The modes, which were grouped under the name "church modes", are the following: *Ionian, Dorian, Phrygian, Lydian, Myxolidian, Aeolian* and *Locrian* [Vallely '99]. All these modes produce a scale based on a sequence of five tones and two semitones. The standard major and natural minor scales in western music are two of the church modes: the Ionian and Aeolian, which correspond to the major and minor scale respectively. As contrast, Irish traditional music uses four of the seven church modes: Ionian (major scale), Dorian, Aeolian (minor scale) and Myxolidian. A list of the most commonly utilised modes by the flute, tin whistle and the uilleann pipe in Irish traditional music is given in [Larsen '03]. The same modes are repeated in Table 2-1, where *M** denotes that the mode *M* is less used than the rest of the modes of the list.

| **Mode tonal centre** | **Mode type** |
|---|---|
| Ionian | D, G and A* |
| Mixolydian | D, G and A |
| Dorian | E, A and B* |
| Aeolian | E, A and B |

**Table 2-1: Most common used modes by the tin whistle, flute and the uilleann pipe (adapted from [Larsen '03])**

It should also be noted that the final note on which the phrases end is usually the tonal centre of the mode [James '02].

### 2.2.2 Forms

There are many different forms of Irish traditional music: singing music as the *sean nós* (which is an old style of singing in the Irish language), dancing music, and non-dance music such as airs. Dancing music represents the majority of the tunes commonly played by traditional musicians, and the most common types are *double jigs*, *hornpipes* and *reels* [Larsen '03]. These types differ in the time signature, tempo, meter and also in the beats where the stress is accentuated. As an example, even though reels and hornpipes can be written in 4/4, hornpipes have a slower pace. In addition, by using 8th notes the first and fifth beats of the bar are more accentuated than in the reels as opposed to the third and seven beats of the bar, which are less accentuated than in the reels [McQuaid '05]. In Table 2-2, a classification of the different types of dance music according to the time signature and meter is illustrated [Larsen '03]

| *Meter* | *Tune Types* | *Time Signature* |
|---|---|---|
| Simple Duple Meter | Reel | 2/2 or 4/4 |
| | Polka | 2/4 |
| | Hornpipe | 2/2 or 4/4 |
| | March | 2/2 or 4/4 |
| | Schottische, Highland, Fling, Highland Fling, | 4/4 |
| | German, Barn Dance | 4/4 |
| | Strathspey | 4/4 |
| Compound Duple Meter | Double jig | 6/8 |
| | Single jig | 6/8 |
| | Slide | 12/8 or 6/8 |
| | March | 6/8 or 12/8 |
| Simple Triple Meter | Waltz | 3/4 |
| | Mazurka, Varsovienne | 3/4 |
| Compound Triple Meter | Slip jig | 9/8 |

**Table 2-2: Types of dance music (adapted from [Larsen '03])**

Irish traditional music is normally played in unison, a technique in which all the instruments play either at the same pitch or at the octave (or double octave) above or below [Prout '06]. In recent years, it has also seen the incorporation of simple harmonic accompaniment [Carolan '06].

### 2.2.3 Ornamentation

Ornamentation plays a very important role in Irish traditional music, and it is used for giving more expression to the music by altering or embellishing small pieces of a melody. However, it is understood differently to classical music, which adds music expression by adding notes to the melody. By contrast, the ornamentations in traditional music are part of the note they ornament, being an integral part of the note onset [Larsen '03]. Ornamentation in Irish traditional music is an improvised element of style. Its spontaneous expressivity can only be completely fulfilled in solo performances, as opposed to group playing in which the freedom of the player is restricted by the unison structure of the playing [O'Canainn '93].

There is a considerable diversity in styles of playing ornamentation within Irish traditional music. Ornamentation has been passed between generations by listening and imitation. This explains the lack of agreement in the manner of describing ornamentation, and also the reason why ornamentation notation symbols are not included in the few available transcribed tunes. The players have learned the ornaments by ear and unconsciously adapt them to their personal style.

In [Larsen '03], a pioneering research in the subject of ornamentation is provided. As [Larsen '03] states: "I believe there is no book before this one that has examined the range of ornamentation that exists in Irish traditional music". [Larsen '03] undertook an analytical research of ornamentation techniques from different players. He invented a novel ornamentation notation, transcribing ornamentations never described before. In addition, [Larsen '03] provides a unique set of tunes with the corresponding ornamentation transcription.

However, within this lack of consensus of existing types of ornamentation, four ornaments appear in all the sources among the most common types in Irish traditional music: cuts, strikes/taps, rolls and cranns [Carolan, Duggan, Larsen '03, O'Canainn '93, Vallely '99]. Other ornaments include trills, triplets and slides. The vibrato effect is only used in non-dance tunes such as slow airs. A description of the main ornaments related to the Irish tin whistle is given in Section 2.3. These techniques are also fully applicable to the flute and the uilleann pipe.

## 2.3 Irish tin whistle

Use of the tin whistle dates from the third century A.D. [Mc Cullough '87]. However, it was not until the 1960′s that the instrument started to occupy the important role in Irish traditional music that it has today. Its versatility allows the instrument to be used either by beginners as an introduction to music, or by experienced players in the most sophisticated tunes. This musical flexibility combined with its easy construction, small size and low cost, makes the tin whistle the most popular instrument in today's Irish traditional music [Vallely '99]. Tin whistles come in a variety of different keys. However, the most common is the small D whistle (Figure 2-3), which can be played in the majority of Irish traditional tunes [Larsen '03].

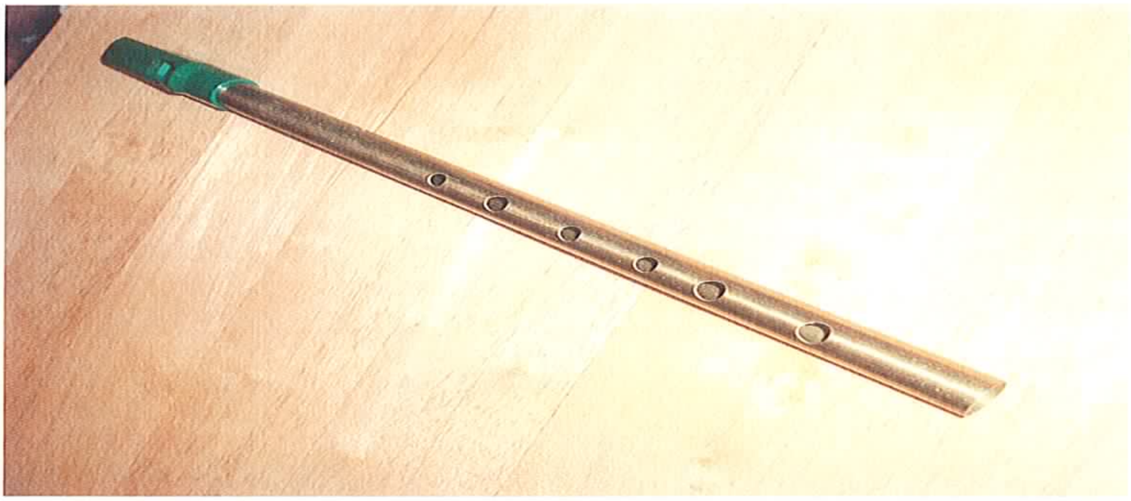

**Figure 2-3: D key tin whistle**

This whistle is a "transposing instrument", which means that when it is played, the note that is heard differs from the written musical notation. For example, for the small D whistle, if a $D_4$ note is written on the score, a $D_5$ note sounds (one octave higher). The small D key whistle is capable of playing in many different modes. Some of them require a half hole covering, which is not practical in many musical situations. Without half covering, the following modes that are very common in Irish traditional Music can be played with the small D Whistle [Larsen '03]:

| Mode tonal centre | Mode type |
|---|---|
| Ionian | D and G |
| Mixolydian | D and A |
| Dorian | E and A |
| Aeolian | E and B |

**Table 2-3: Most played modes by the D key tin whistle (adapted from [Larsen '03])**

The Irish tin whistle is a good example of a slow onset instrument. The accurate detection of its onset is the research topic of Chapter 6. By regulating the air flow, changing the position of the mouth in the tin whistle lip, or altering the position of the fingers that cover the tin whistle holes, tin whistle players can produce strong amplitude and frequency modulations. In addition, successive notes are often played without any intervals by using a legato technique, where the articulation only occurs in the first note of the group. In this case, the energy increase in the slurred notes can be small, or even nonexistent.

### 2.3.1 Acoustic Properties of the Tin Whistle

The tin whistle produces sound by directing air through a channel against a sharp lip, which splits the air stream causing it to vibrate. This instrument acts acoustically as an open pipe at both ends. In this case, as illustrated in Figure 2-4, the pressure component is zero at both ends, which are called pressure nodes [Howard '01].

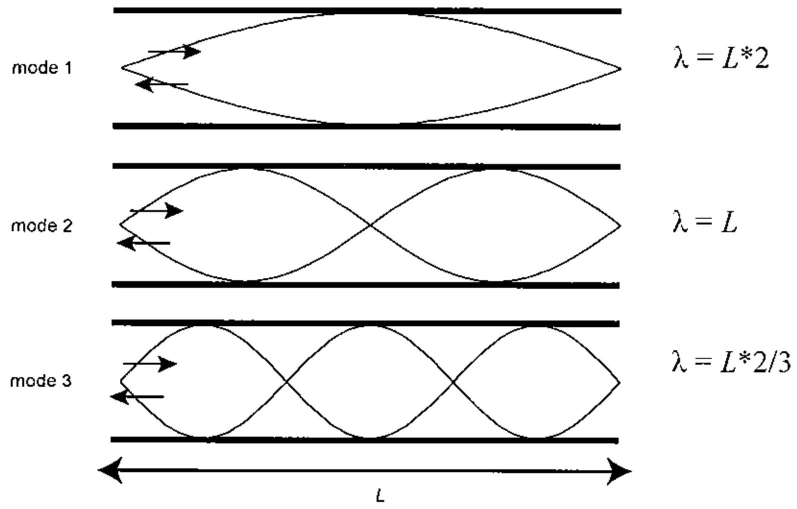

**Figure 2-4: Acoustical pressure of the first three modes of an open pipe at both ends. The left and right y labels denote the mode number and the λ of the standing wave respectively**

As can be seen in the top row of Figure 2-4, the minimum frequency for a standing wave that produces pressure nodes at the ends of a pipe of length $L$ occurs at half the wavelength λ [Howard '01]. This frequency, $f_{low}$, corresponds to the lowest note that the tin whistle can play, which is given by:

$$f_{low} = \frac{v}{\lambda} = \frac{v}{2L} \qquad (1)$$

where $v$ is the velocity of sound

Equation (1) shows that the frequency of the standing wave depends on the length of the tube $L$. In the case of the tin whistle, the length $L$ of Equation (1) is reduced by the action of lifting the fingers off the holes. This has the result of modifying the value of $f_{low}$ and consequently the note within the same register of the instrument. By closing all the holes, the lowest note of the instrument is played (e.g., D for a D key tin whistle). Then, lifting all the fingers off the holes successively from the bottom end provides a diatonic scale (E, F#, G, A, B and C for a D key tin whistle).

As can be seen in the bottom and middle plot of Figure 2-4, any multiple of half wavelengths will also fulfil the condition of having a pressure node at both ends. Consequently, standing waves will also fit between the pipe ends if the frequency $f_n$ is given by:

$$f_n = nf_{low} \qquad (2)$$

where $n$ denotes the mode number

In the tin whistle, the utilised note range covers two octaves. In order to play in the second octave, the player needs to overblow to reach the second mode. This produces a standing wave with frequency $f_n$=2* $f_{low}$ (see Equation (2)), which is depicted in the middle plot of Figure 2-4. Then, the same fingering as in the low octave can be utilised by the tin whistle player.

Each of the standing waves depicted in Figure 2-4 corresponds to a sine frequency component. However, when playing a note with frequency $f_{low}$ other harmonics with frequency $fn$ are also present (where $n$ in $fn$ denotes the harmonic number) [Fletcher '98]. The spectrum that contains the note harmonics varies depending on the note played, the loudness utilised by the player, the instrument manufacturer and the recording conditions. However, the tin whistle generally produces a prominent fundamental harmonic in relation to the other harmonics. In addition, the blowing pressure utilised by the player increases with the note played [Martin '94]. This is illustrated in Figure 2-5, where the magnitude spectrum of notes E4, A4 and E5 played by the same player utilising a Copeland D key tin whistle are shown. The fundamental harmonic is prominent in the three cases and its magnitude value increases with the note frequency. Some manufacturers produce tin whistles whose third harmonic can have a stronger amplitude value than the second. This phenomenon can be seen in E4 and A4 spectrum of Figure 2-5.

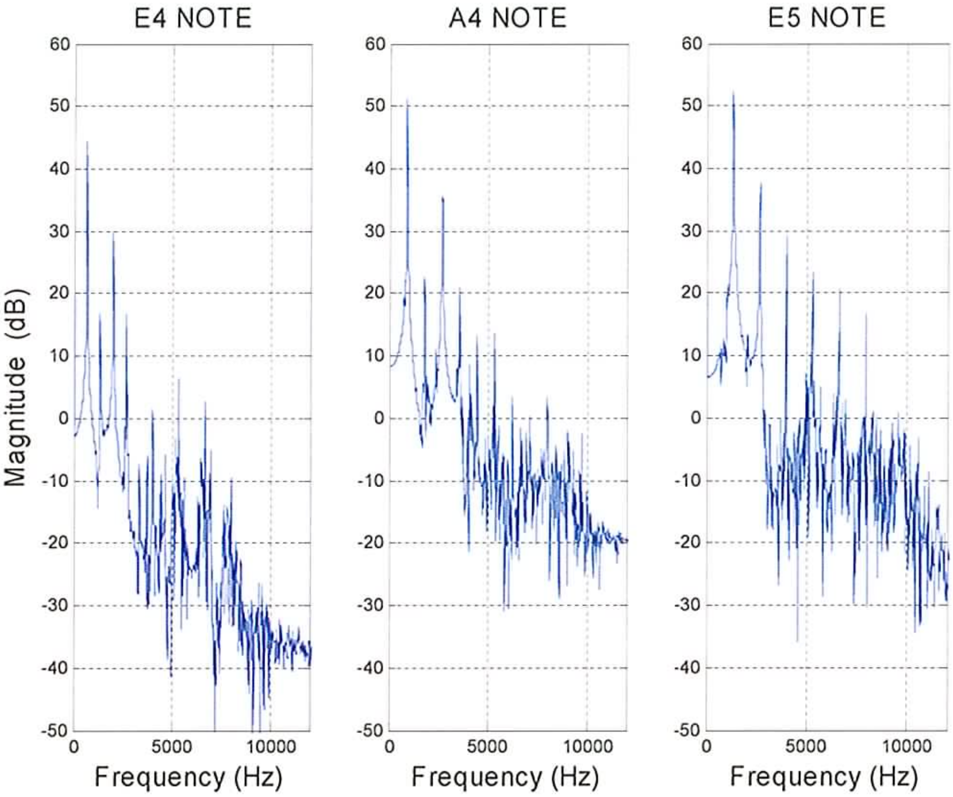

**Figure 2-5: Magnitude spectrum of three different notes (E4, A4 and E5) played by the tin whistle**

### 2.3.2 Ornamentation

It is widespread among tin whistle players to utilise ornamentation to embellish the notes played. As previously seen in Section 2.2, there are many different types of ornamentation in Irish traditional music. [Larsen '03] splits them into two categories: single-note ornaments and multi-note ornaments. Thus, the most frequently used ornaments mentioned in Section 2.2 (cuts, strikes, rolls and cranns) are classified and defined as follows:

- **Single-note ornaments: cuts and strikes**
  - The **cut** is a subtle and quick lift of the finger covering its hole followed by an immediate replacement, which increases the pitch [Larsen '03]. This type is by a large extent the most commonly single-note ornament used in Irish traditional music.
  - The **strike** is a rapid impact of an uncovered hole that momentarily lowers the pitch.

Even though cuts and strikes have a pitch, their sound duration is very brief, and not perceived as having a discernible pitch, note or duration on their own [Larsen '03]. Therefore, as opposed to classical music they can not be considered notes, nor graces notes, but rather are just part of the onset of the note that they ornament [Larsen '03].

- **Multi-note ornaments: rolls and cranns**

By playing certain combinations of single-note ornaments, multi-note ornaments can be built. The most common types are introduced as follows:

- The **long roll** is a group of three slurred notes of the same pitch, where the second and third notes are played with a cut and a strike respectively. This roll is the most played multi-note ornament. There is also a **short roll** version, which is like the long roll but without the first unornamented note [Vallely '99].
- The **long crann** only uses cuts to form the multi-note ornament. It is comprised of three slurred notes of the same pitch, where all the notes are ornamented with a cut except the first note [Duggan]. This ornament was created by the uileann pipe in order to accomplish a staccato "roll" in the lowest note D of the chanter [Vallely '99]. The same principle applies to the simple-system flute and the D tin whistle, where it is not possible to strike a D note, so a cut is used instead. The **short crann** version is like the long crann but removes the first unornamented note.

The most common use of cuts and strikes is to separate two notes of the same pitch [Vallely '99]. However, they are also used to ornament ascending or descending independent notes in the melody [Larsen '03], as in short versions of multi-note ornaments.

The long versions of the multi-note ornaments are more frequently played than the short versions. In contrast to short versions, the first note of the long versions is always unornamented. This provides more time for the player to prepare the ornaments that separate repeating notes [Larsen '03].

- **Other ornaments:**
  - **Slides**: inflection of the pitch (generally upwards), which can also occur outside the onset part of the note [Larsen '03].
  - **Triplets**: a rising or descending sequence of three notes played in the same time than two notes [Duggan].

Other articulations such as throating and tonguing can also be played by the tin whistle. However, as opposed to the above listed ornaments these articulations do not have an implicit pitch, and are not perceived as ornamental [Larsen '03].

## 2.4 Discussion

The main instruments that are commonly played in Irish traditional music have been described in Section 2.1. Even though the range is very wide, six instruments are encountered in the majority of the tunes: the fiddle, the uilleann pipe, the simple-system flute, the concertina the button accordion and especially the tin whistle, which currently play a very important role. The most widely played tin whistle is the D key tin whistle, and a description of its characteristics has been given in Section 2.3.

Table 2-1 illustrates the most common played modes in Irish traditional music for the Irish tin whistle, flute and uilleann pipe. Since these three instruments dominate the majority of the tunes, Table 2-1's modes will also be frequently used by the other instruments. In Table 2-2, the modes that the Irish tin whistle can play without using half-covering are shown. By comparing Table 2-1 and Table 2-2, it can be derived that apart from G Mixolydian and A Aeolian, the most commonly utilised modes of Table 2-1 can be played without using half-covering in a D key tin whistle. If a tune is played in G

Mixolydian and A Aeolian, the player can change to a tin whistle that can handle F naturals such as the C key Whistle. In these cases, the tunes are played as if they were in A Mixolydian or B Aeolian (respectively) one note higher than written in the score, without using half-covering. As an example, an F note in G Mixolydian is turned into a G note in A Mixolydian. Then, by playing with a C key tin whistle as if using a D key tin whistle, an F note will sound.

Therefore, by using a D key tin whistle the majority of the most common modes of Irish traditional music can be played without using half covering. If a tune is in a different mode, the player can change to a C key tin whistle to keep playing without using half covering.

Ornamentation in Irish traditional music is of great importance. In Section 2.2, the most commonly occurring types of ornamentation have been introduced: cuts, strikes/taps, rolls and cranns. The uilleann pipe established an important legacy of ornamentation techniques within Irish traditional music. Since the tin whistle belongs to the same family of wind instruments, the tin whistle ornamentation techniques introduced in Section 2.3 will also be an illustrative example of ornamentation within Irish traditional music.

## 2.5 Conclusions

A description of the different instruments within Irish traditional music has been given in Section 2.1. Then, the main aspects of the structure of the Irish traditional music have been introduced in Section 2.2. Section 2.3 focuses on the musical features of the Irish tin whistle. The discussion of Section 2.4 shows that the D key Irish tin whistle illustrates well the features of Irish traditional music. This is reflected in the great importance that the instrument has within Irish traditional music, its ornamentation techniques, and the modes that can play even without using half covering.

In Chapter 6, an onset detector based on the characteristics of the Irish tin whistle is presented. In addition, a novel transcription algorithm for detecting ornamentation played by the Irish tin whistle is presented in Chapter 7.

In the next chapter, an introduction to comb filter techniques is given. Comb filter methods have been used in the presented thesis to implement applications within an Irish traditional context. The applications are introduced in Chapters 8 and 9, which present an onset detector and a pitch detector respectively.

# 3 Comb filtering

Comb filters are so named because the peaks and notches in their magnitude frequency response resemble the teeth of a comb (Figure 3-2). This harmonic spectral shape can be interpreted as a bank of band pass filters (BPF) equally spaced over the frequency axis. Comb filter techniques are widely utilised in many musical applications, for example in delay based audio effects [Fernández-Cid '98], sound reinforcement [Nehorai '86], reverberation techniques [Moorer '85], music separation [Gainza '04b, Miwa '99b], onset detection [Gainza '05b], and pitch detection [Gainza '05a, Moorer '74, Tadokoro '03]. In this Chapter, a summarised description of the main comb filter techniques is given. Section 3.1 focuses on FIR comb filter methods, which includes its standard structure, the parallel configuration version and interpolation techniques. Following this, a description of the characteristics of IIR comb filter techniques is introduced in Section 3.2. Finally, a discussion of the different techniques and some conclusions are given in Sections 3.3 and 3.4

## 3.1 FIR Comb Filters

### 3.1.1 Standard structure

By using FIR comb filters, the comb spectral shape can be obtained by summing a discrete input signal $x$ with a delayed version of the same discrete signal $x$. The FIR comb filter block diagram is depicted in Figure 3-1, and the difference equation and transfer function are represented as follows [Moorer '74, Tadokoro '03]:

$$y[n] = x[n] + b_1 * x[n-D] \tag{3}$$

$$H(z) = 1 + b_1 * z^{-D} \tag{4}$$

where $b_1$ is a factor which scales the gain of the filter between 1+ $b_1$ and 1- $b_1$, and $D$ is the delay in samples.

The unit sample response $b$ is then built as follows:

$$b = [1 \; numZeros \; b_1] \tag{5}$$

where *numZeros* is a vector of $D$-1 zeros

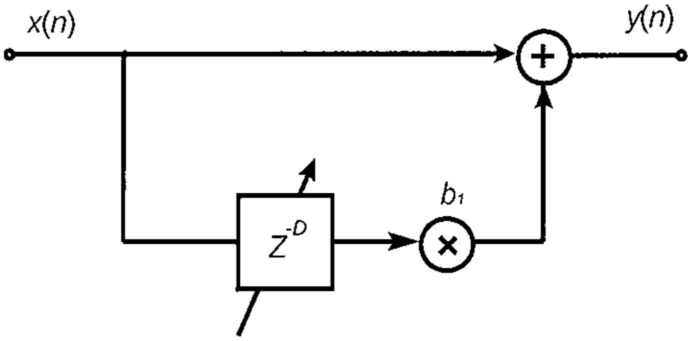

**Figure 3-1: FIR comb filter block diagram**

The comb effect results from phase cancellation and summation between the delayed and original signals. This can be seen in Figure 3-2, where the magnitude response of two filters with $b_1$ = 1 (dashed line) and $b_1$ = -1 (solid line) respectively is depicted. The sampling rate $f_s$ is 44100 Hz and the delay $D$ is 4 for both filters.

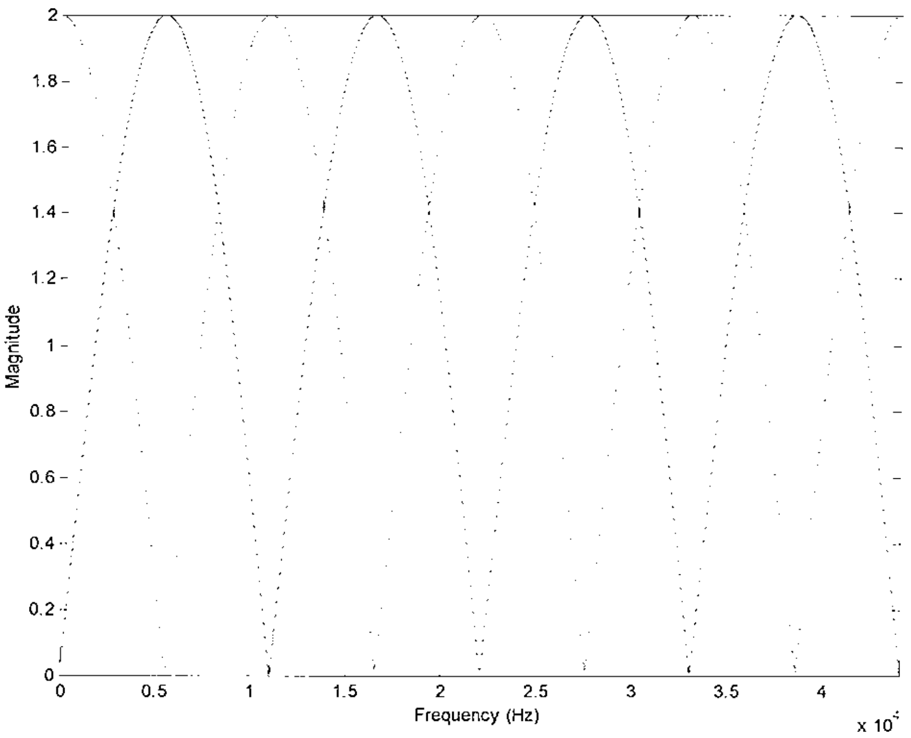

**Figure 3-2: FIR comb filter magnitude response using $b_l$=1 (dashed line) and $b_l$=-1 (solid line)**

From Figure 3-2, it is apparent that at frequencies $n*(f_s/D) \leq f_s$, where $n$ is an integer, the delay $D$ causes a 360 degree shift between the original and delayed signal producing summation and cancellation in the filter with $b_l = 1$ and $b_l = -1$ respectively, which produces peaks and notches in the filter magnitude frequency response at frequencies: 11025, 22050, 33075 and 44100 Hz.

### 3.1.2 Parallel FIR comb filters structure

Narrower comb filter widths and larger gains can be obtained by connecting $N$ delay lines in cascade. The block diagram is depicted in Figure 3-3, and the difference equation and transfer function are represented as follows [Bernhardt '74, Proakis '95]:

$$y[n] = x[n] + b_1 * x[n - D] + \ldots + b_N * x[n - D * N] \tag{6}$$

$$H(z) = 1 + \sum_{k=1}^{N} b_k * z^{-D*k} = \sum_{k=0}^{N} b_k * z^{-D*k} \tag{7}$$

where $b_0 = 1$

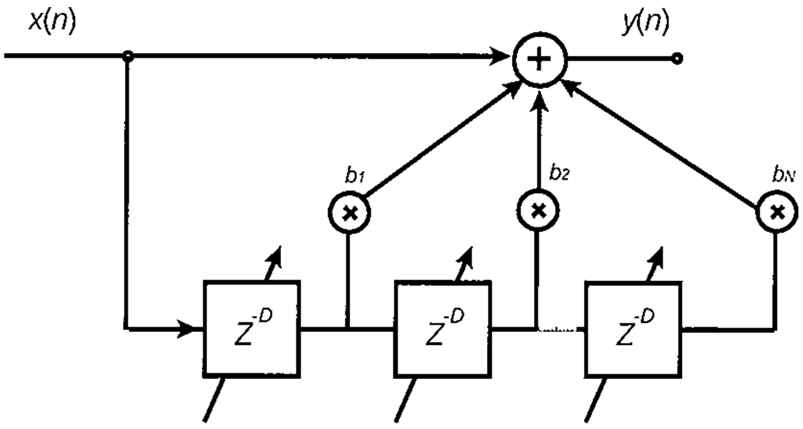

**Figure 3-3: Parallel FIR comb filters structure**

By setting all vector coefficients to $b_k = 1$, where $1 <= k <= N$, the gain varies from 1 - $b_1$ to 1+ $b_1 * N$, and Equation (7) becomes [Bernhardt '74, Proakis '95]:

$$\sum_{k=0}^{N} z^{-D*k} = \frac{1 - Z^{-D(N+1)}}{1 - Z^{-D}} \tag{8}$$

Thus, the frequency response is given by [Bernhardt '74, Proakis '95]:

$$H(f) = \frac{\sin[\pi f D(N+1)]}{\sin(\pi f D)} \cdot e^{-j\pi fDN} \tag{9}$$

Equation (9) produces zeros at frequency multiples of $f = 1/((N+1)*D)$, except for multiples of $1/D$, where $H(f)$ becomes equal to $N$+1. The width of the lobes is inversely proportional to the number of $N$ filters connected. This is illustrated in Figure 3-4, where the magnitude response of a parallel structure with $N = 5$ and $N = 8$ filters is depicted in the left and right row respectively. The comb filters were built with a delay $D = 127$

samples, $b_k = 1$ and $fs$ equal to 44100 Hz. Thus, both magnitude responses produce peaks at multiples of $fs/D \approx 347$ Hz. In between those peaks, zeros occur at multiples of $fs/(D*(N+1))$, which correspond to multiples of ≈ 57.8 and ≈ 30.5 Hz for the case of an $N = 5$ and $N = 8$ filters structure respectively.

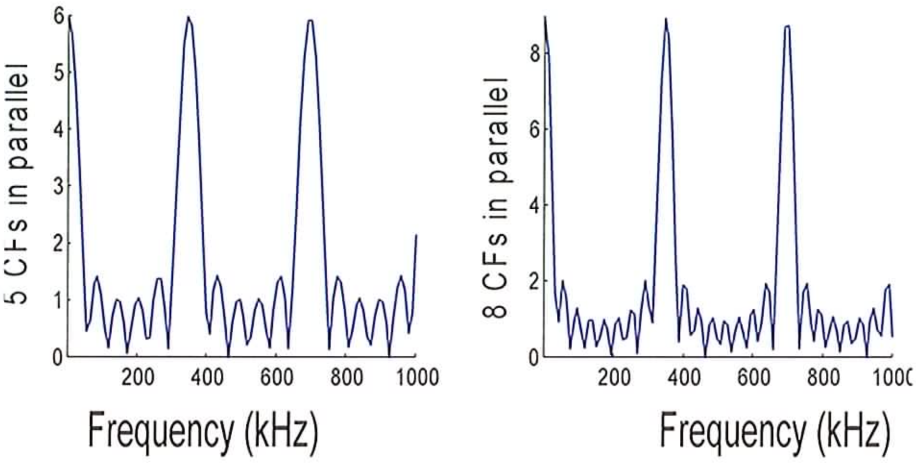

**Figure 3-4: Parallel comb filter magnitude responses using $N = 5$ (left plot) and $N = 8$ (right plot)**

### 3.1.3 Using interpolation techniques

FIR Comb Filters can also be built by the use of interpolation techniques [Proakis '95]. The technique is explained as follows:

The $Z$ transform of an FIR filter with impulse response $h(n)$ with length $M$ is given by:

$$H(z) = \sum_{n=0}^{M} h(n)Z^{-n} \tag{10}$$

Replacing $Z$ by $Z^D$, the above equation becomes:

$$H'(z) = \sum_{n=0}^{M} h(n)Z^{-Dn} \tag{11}$$

which is equivalent to interpolating $D$ zeros in between the $h(n)$ samples.

Thus, the frequency response of Equation (11) corresponds to:

$$H'(w) = \sum_{n=0}^{M} h(n)e^{-jnDw} = H(Dw) \qquad (12)$$

where $H(w)$ is the frequency response of Equation (10).

Thus, by interpolating $D$ zeros in between the $h(n)$ samples, $H(w)$ is repeated $D$ times in between 0 and $2\pi$.

## 3.2 IIR Comb Filters

High amplitude gains and narrow peaks can be obtained in fewer operations by using IIR Comb filters. The IIR comb filter block diagram is depicted in Figure 3-5, and the difference equation and transfer function are represented as follows [Dutilleux '98]:

$$y[n] = x[n] + b_1 * y[n-D] \qquad (13)$$

$$H(z) = \frac{1}{1 - b_1 * z^{-D}} \qquad (14)$$

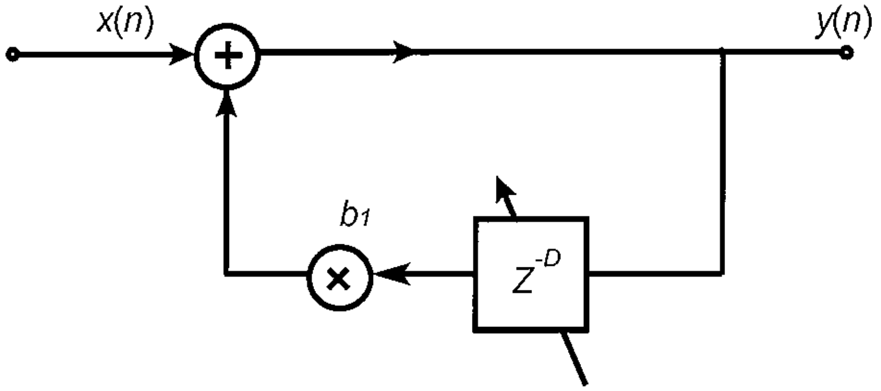

**Figure 3-5: IIR comb filter block diagram**

To ensure stability, $b_1$ must be $\leq 1$. The gain varies from $1/(1+ b_1)$ to $1/(1- b_1)$, and the peaks become narrower as $b_1$ gets closer to 1. Due to its recursive structure, at each cycle around the feedback loop, the input signal is scaled again by $b_1$, so that after $p$ cycles the signal has been multiplied by $b_1{}^p$[Dutilleux '98]. Comparing Figure 3-2 and Figure 3-6, it can be appreciated that IIR comb filters produce narrower peaks and flatter notches at the same frequencies as for FIR comb filters.

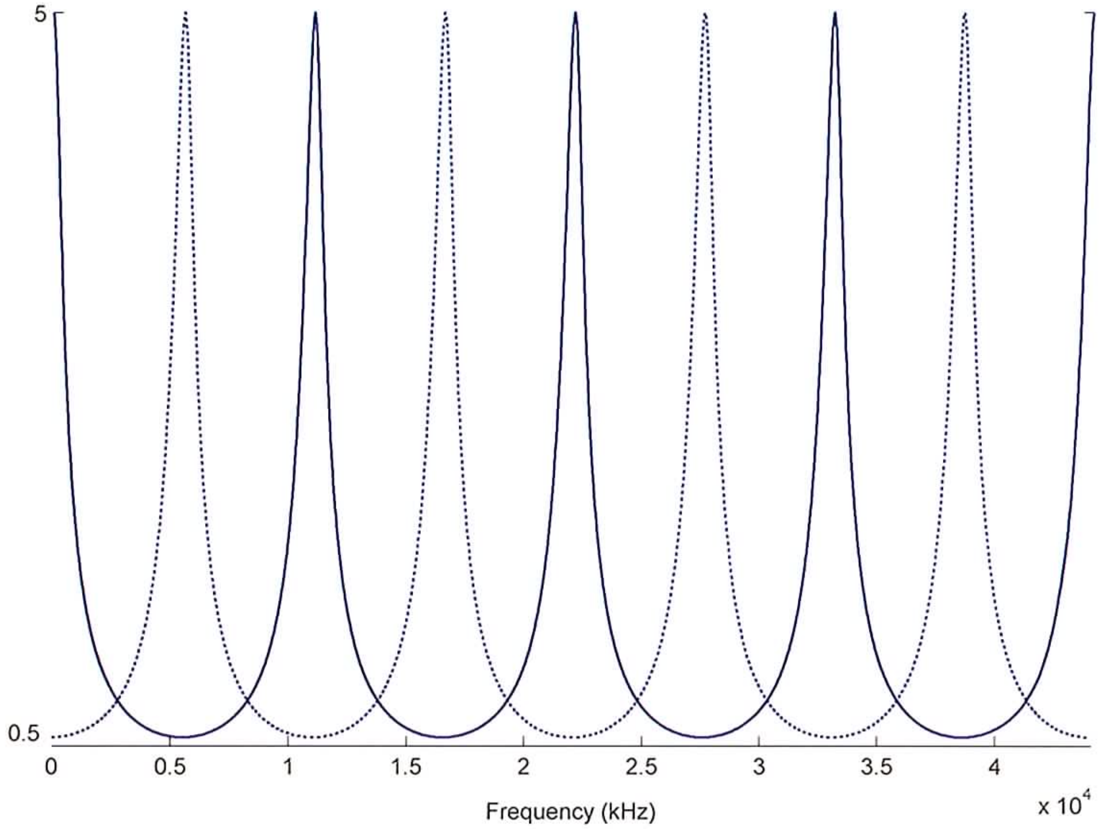

**Figure 3-6: IIR comb filter magnitude response using $b_1 = 0.99$ (solid line) and $b_1$=-0.99 (dashed line)**

## 3.3 Discussion

In this chapter, different techniques to build comb filters have been introduced. Comb filters are characterised by a magnitude response having peaks or nulls equally spaced over the frequency axis. This can be accomplished by using FIR comb filters, which only require two parameters: the delay filter $D$ and the gain $b_I$. Another advantage of using FIR comb filters is that since the structure only uses delays, the filters can be efficiently computed in the time domain.

If narrower comb filter widths and larger gains are required, a parallel comb filter structure can be used. In this case, the width of the lobes in the magnitude response gets narrower with the number of $N$ filters connected in parallel. However, apart from being computationally more intense, other nulls and peaks arise in between the main peaks of the magnitude response.

High amplitude gains and narrow filter widths can also be obtained using IIR Comb filters. By comparing Figure 3-2 and Figure 3-6, it can be appreciated that IIR comb filters produce narrower peaks and flatter notches at the same frequencies as the FIR comb filters. However, due to its recursive nature the stability of the filter has to be ensured. In addition, the phase of the filter will not be linear.

Another manner of designing FIR comb filters is by using interpolation techniques, which repeats the magnitude response of a filter with impulse response $h(n)$, $D$ times along the frequency axis. This filter has the advantage that by knowing $h(n)$, comb filters with interesting magnitude response shapes can be designed. However, long $h(n)$ will also produce very long comb filter lengths.

## 3.4 Conclusions

Different techniques to construct comb filters have been introduced in this chapter. The methods include standard FIR comb filters, parallel FIR comb filters, interpolation techniques and IIR comb filters.

The use of comb filter methods has been significantly exploited in this thesis. The harmonic shape magnitude response that can be attained by using comb filters is a very useful feature when dealing with harmonic signals. This is usually the case in Irish traditional music, where percussion is rarely utilised. Based on this, several musical applications within Irish traditional music have been developed. In Chapter 9, the use of IIR comb filters is investigated to build a multi-pitch estimator. FIR comb filter techniques form the basis of the onset detector presented in Chapter 8. This is related to the topic of the next chapter, which provides a description of the different existing approaches to perform onset detection.

# 4 Onset Detection

## 4.1 Introduction

A musical onset is defined as the precise time when a new note is produced by an instrument [Bello '04]. The onset of a note is very important in instrument recognition, as the timbre of a note with its onset removed can be very difficult to recognise [Grey '75]. Masri states that in traditional instruments, an onset is the stage during which resonances are built up, before the steady state of the signal [Masri '96b]. Other applications use separate onset detectors in their systems. For example, in rhythm and beat tracking systems the tempo is obtained by calculating the frequency of the onset occurrence [Scheirer '98]. Music transcriptors utilise an onset detector to obtain the exact time when the detected new note has arisen [Gainza '04a, Gainza '04c, Klapuri '01, Klapuri '03 , Marolt '02]. Time stretching algorithms use onset detectors to time scale the steady part of the signal whilst leaving the onset part unaltered [Dorran '04]. In [Virtanen '00], a music instrument separator utilises onset information to assign harmonics with common onset to the same source.

The two most common types of onsets can be classified as follows:

- A **sharp onset**, which is a short duration of the signal with an abrupt change in the energy profile [Masri '96b], appearing as a wide band noise burst in the spectrogram (Figure 4-1). This change manifests itself particularly at high frequencies and is typical in percussive instruments.

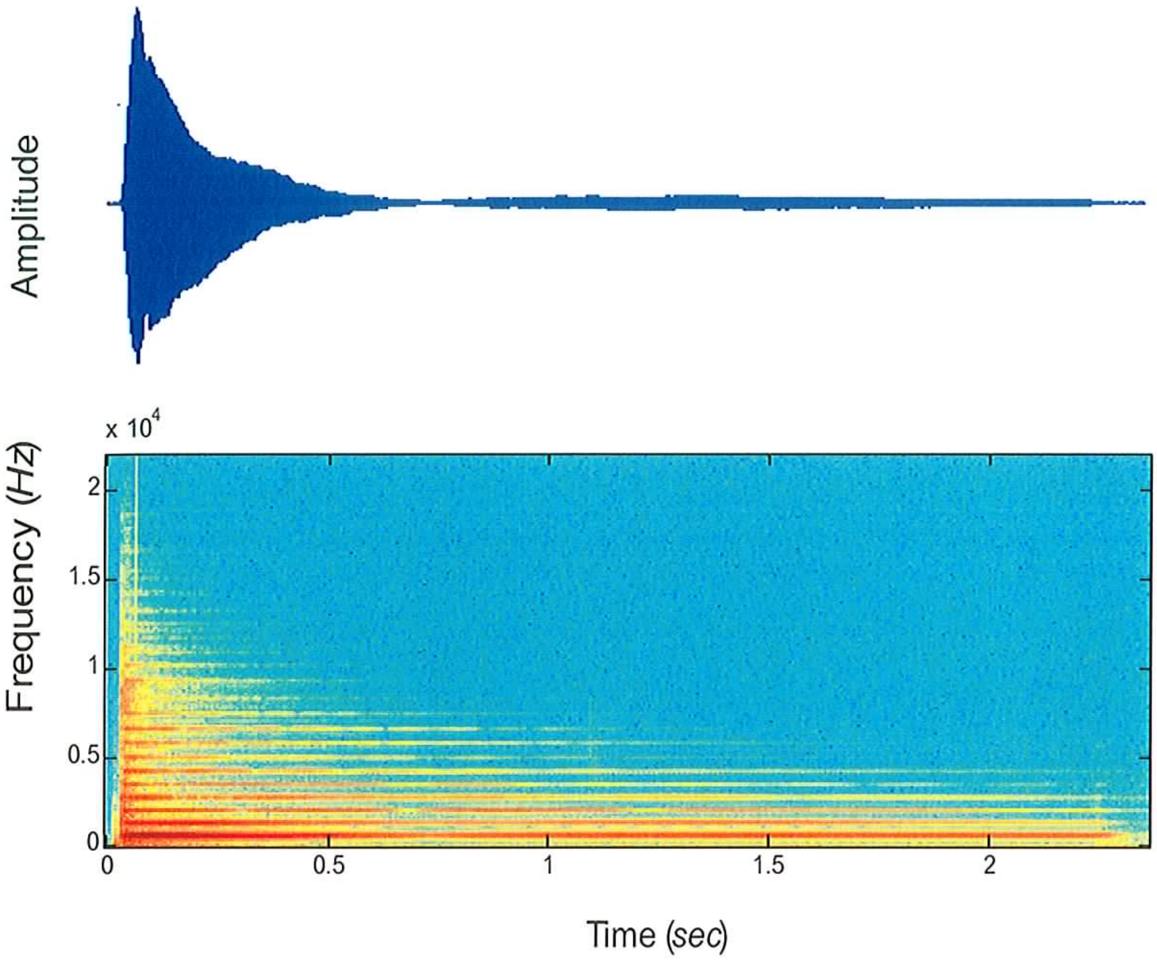

**Figure 4-1: Example of a spectrogram of a piano playing $C_4$**

- A **slow onset**, which typically occurs in wind instruments like the flute or the tin whistle, is more difficult to detect. As illustrated in Figure 4-2, the onset takes a much longer time to reach the maximum onset amplitude value and has no noticeable change at high frequencies [Duxbury '02].

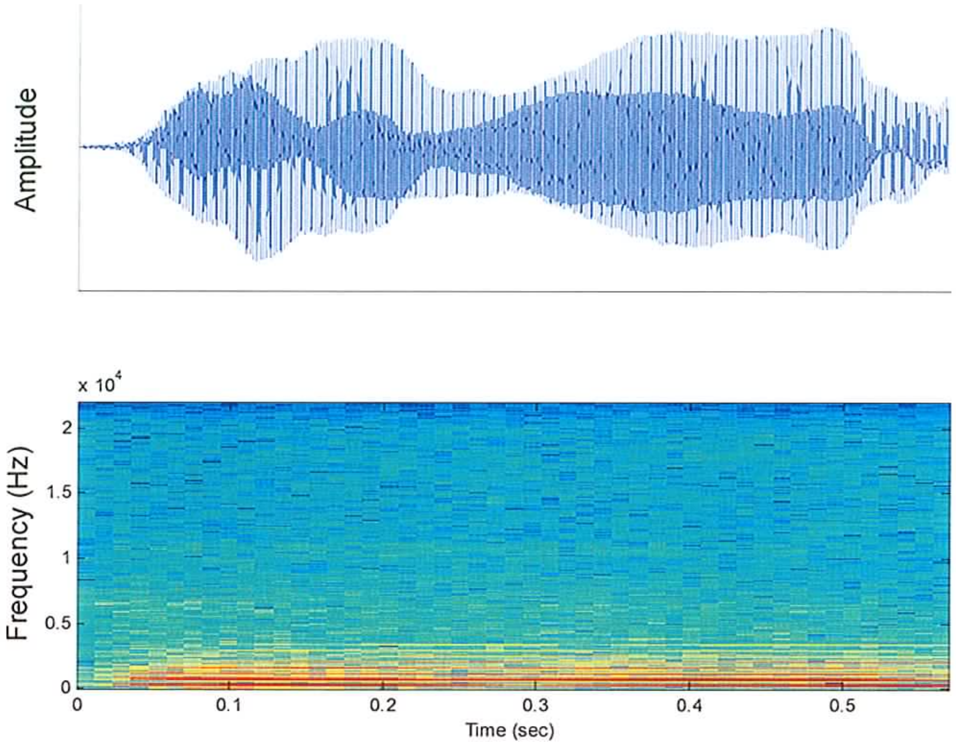

**Figure 4-2: Example of a spectrogram of a flute playing A4**

In this review chapter, the main existing onset detection methods are investigated. The methods are classified by the signal information they utilise to detect its onset part: energy and/or phase. The advantages and disadvantages of the existing approaches to detect sharp or slow onsets are also discussed in Section 4.5, which leads to the conclusions described in Section 4.6, and to the implementation of effective onset detectors introduced in Chapters 6 and 8.

## 4.2 Energy based onset detection approaches

Early work which dealt with the onset detection problem analyses the amplitude envelope of the entire input signal for the purpose of onset detection [Chafe '86]. However, this approach only works for signals that have a very prominent onset. The same limitations apply to the standard short time energy *SE*, which is given by:

$$SE(n) = \sum_{m=(n-1)h}^{nh} (x[m])^2 \qquad (15)$$

where $n$ is the hop number, $h$ the hop size and $x$ a windowed input signal.

Multi-band approaches provide information on specific frequency regions where the onset occurs. This was first suggested by Bilmes [Bilmes '93], who computed the short time energy of a high frequency band. Then, the system computes the slopes of the energy over time searching for a value that reaches a given threshold. The onset time will be the maximum slope in a predefined region once a given threshold has been attained.

In order to synchronise the analysis window of the Deterministic Plus Stochastic model developed by [Serra '89] to the transient events, [Masri '96b] proposes a method for detecting sharp onsets. This frequency domain method gives more weight to the high frequency content (HFC) of the signal using the following function:

$$HFC = \sum_{k=2}^{\frac{N}{2}+1} \left\{ |X(k)|^2 * k \right\} \qquad (16)$$

where $N$ is the FFT array length, $X(k)$ the FFT $k_{th}$ bin and $\frac{N}{2}+1$ corresponds to the Nyquist frequency.

The condition for onset detection is that the rise in *HFC* between two consecutive frames multiplied by the normalised *HFC* of the current frame has to be greater than a given threshold $T_D$:

$$\frac{HFC_n}{HFC_{n-1}} * \frac{HFC_n}{E_n} > T_D \qquad (17)$$

where $E_n$ *is* the standard energy given by Equation (15).

A problem related to energy methods that give more emphasis to the high frequency bins, is that their performance is highly related to the characteristics of the signal. Considering Figure 4-3, where a piano signal that plays C3, C4, C5 and C6 consecutively is depicted in the top plot. It can be appreciated that by utilising the HFC method and the detection function of Equation (17), which are depicted in the middle top plot and bottom plot respectively, high pitched notes are emphasised. As a reference, the energy function is also depicted in the middle top plot. Noisy artefacts, which normally arise in high frequencies will also be accentuated.

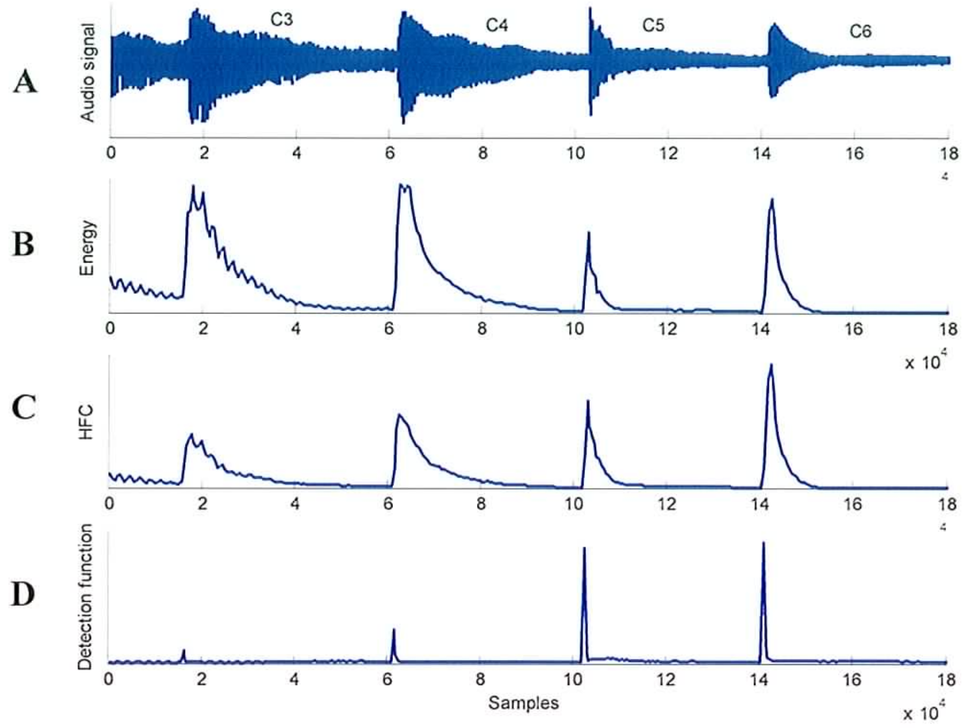

**Figure 4-3: Energy function (plot B), HFC function (plot C) and Masri's detection function (plot D) of a piano signal (plot A).**

A method to calculate the spectral difference between frames is proposed in [Masri '96a], which is an estimation of the average of the increase in energy per frame:

$$dX_n(k) = X(k,nh) - X(k,(n-1)h) \tag{18}$$

The detection function can now be calculated as follows:

$$DM(n) = \sum_{k=1}^{N/2} dX_n(k) \tag{19}$$

This detection function is a more effective method of detecting onsets with weak high energy than the HFC method, since all the frequency bins are equally weighted. An example is depicted in Figure 4-4, where the spectral difference method is applied to a Banjo signal.

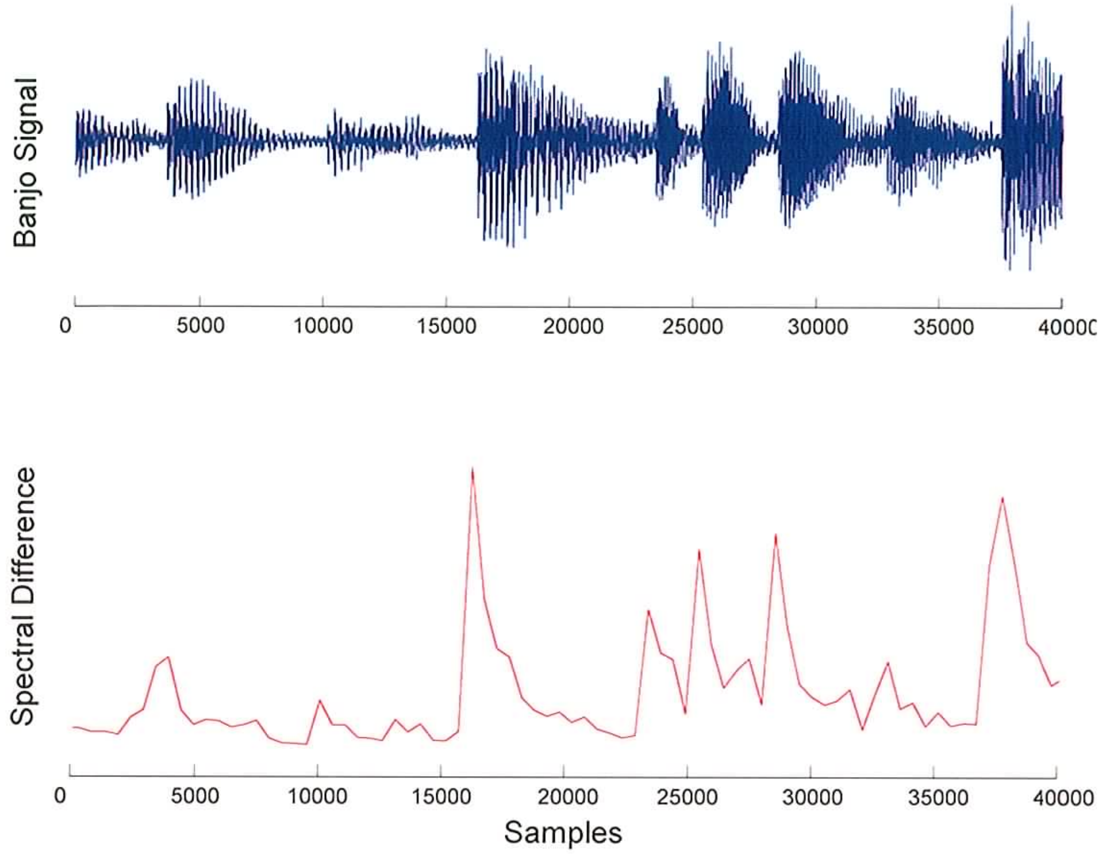

**Figure 4-4: Example of Masri's spectral difference onset detection of a banjo excerpt of "The Cock and the Hen", played by RIRA**

[Scheirer '98] presents a system for estimating the beat and tempo of acoustic signals requiring onset information. A filterbank divides the incoming signal into six frequency bands, each one covering one octave. The amplitude envelope is extracted for each band by rectifying the signal, and then it is smoothed using a 200 ms Half Hanning window (raised cosine) [Proakis '95]. The difference between the Hanning and the Half Hanning window is shown in Figure 4-5, where it can be appreciated that the Half Hanning window (plot D) contains a less sharp stop band than the "full" Hanning window plot C. [Klapuri '98] states that this feature masks fast amplitude modulations but emphasises the most recent inputs. An example of an amplitude envelope signal smoothed by using full and half Hanning windows is depicted in Chapter 6,

Figure 6-4.

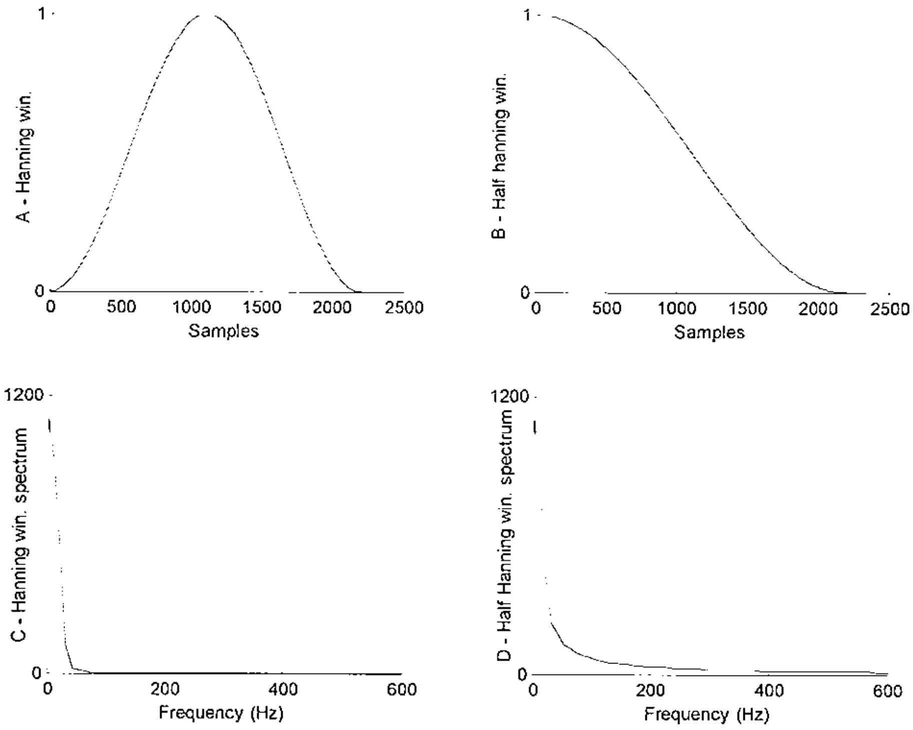

**Figure 4-5 Comparison between Hanning and Half Hanning windows**

After smoothing the signal, the amplitude envelope is decimated to 200 Hz. As an example, the output signal after applying a band pass filter [200Hz - 400Hz] to a violin signal is illustrated in Figure 4-6, plot A. The amplitude envelope and the decimated amplitude envelope of the filter output are depicted in plot B and C respectively. The decimating factor utilised is 44100/200 = 220. Decimation also occurs when analysing the short time energy of a signal, where the signal is decimated by factor $h$, which is the hop size. As a comparison, the energy detection function of the same band filtered violin signal using a $h$ = 200 samples is depicted in plot D. It can be appreciated that the decimated amplitude envelope has the signal shape of an unsmoothed energy envelope.

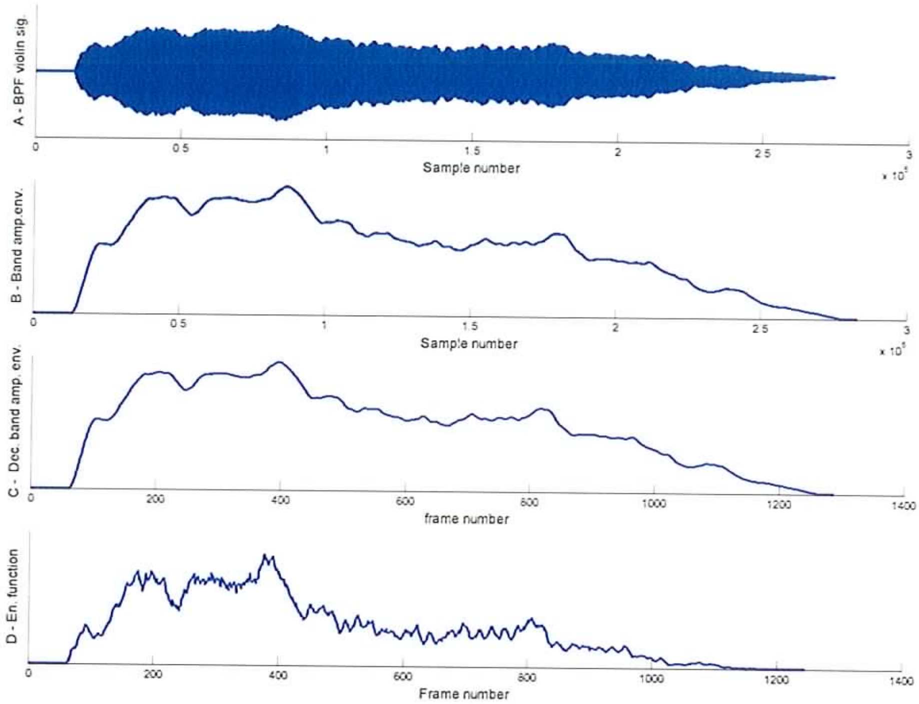

**Figure 4-6: Example of the use of Scheirer's onset function detection**

Next, Scheirer applied the rectified first order difference to the decimated amplitude envelope as follows:

$$D(t) = \frac{d}{dt}(A(t)) \qquad (20)$$

where $A(t)$ is the decimated amplitude envelope.

The equivalent operation of Equation (20) in a STFT analysis will be given by:

$$dE(n) = E(n) - E(n-1) \qquad (21)$$

where $E(n)$ is the energy of the frame $n$.

Even though Scheirer did not attempt to perform specific onset time detection, his model was the basis for future onset detectors.

[Klapuri '98] develops an onset detector system based on Scheirer's model [Scheirer '98]. He also splits the signal into several frequency bands and obtained the amplitude envelope and the first order difference in each band. However, in order to determine the onset times he divides the first order difference by the amplitude envelope obtaining the *relative difference function W(t)* as follows:

$$W(t) = \frac{\frac{d}{dt}(A(t))}{A(t)} = \frac{d}{dt}(\log(A(t))) \qquad (22)$$

where $A(t)$ is the amplitude envelope.

*W(t)* gives a better estimation of onset times of signals that take some time to reach the point of maximum onset slope [Klapuri '98]. The amount of change is related to the absolute signal level, the same amount of increase being more relevant in a quiet signal [Klapuri '98]. The relative difference function *W(t)* and the first order difference function are utilised to obtain the onset candidate time and its corresponding prominence value (magnitude) respectively.

After getting $W(t)$ for each band, only peaks above a given threshold $T_{det}$ are considered as onset candidates. Then, the onset candidate prominence values are calculated at the closest peak in the first derivative function after the onset candidate time. Onset candidates that fall in a 50 ms window of a more intense component are dropped out for every band. Then, the remaining peaks in all bands are sorted in time order, and a new prominence value is given for all peaks, summing the prominence value of onset candidates within a 50 ms time window. Once more, the most intense candidate in a 50 ms time window is kept, and finally only candidates above a second threshold $T_{final}$ are maintained as onsets [Klapuri '98].

The top plot of Figure 4-7 depicts the onset part of the decimated amplitude envelope of Figure 4-6, plot C. The first order difference function of the top plot of Figure 4-7 is obtained by applying Scheirer's system. The result is depicted in the middle plot where it can be appreciated that the maximum in the function is delayed from the real onset time. The bottom plot depicts the relative difference function of the decimated amplitude envelope, where the maximum in the function is closer to the onset time. However, as it can be seen in the first frames of the relative difference function plot of Figure 4-7, a small amplitude increase between small amplitude envelope values can be excessively high in relation to the amplitude envelope, which can cause spurious peaks.

In [Klapuri '99], a psychoacoustical based implementation of [Klapuri '98]'s system is presented. The method utilises a bank of 21 non-overlapping filters covering the critical bands of the human auditory system, and incorporating Moore's psychoacoustic loudness perception model [Moore '97]. To obtain the loudness of every band onset candidate, their corresponding intensities are calculated, which is achieved by multiplying the peak prominence onset candidate value by the band centre frequency. The peaks in all frequency bands are combined as in [Klapuri '98]. Then, the loudness of the onset candidates are obtained by summing the intensity values of the onset candidates within a 50 ms time window.

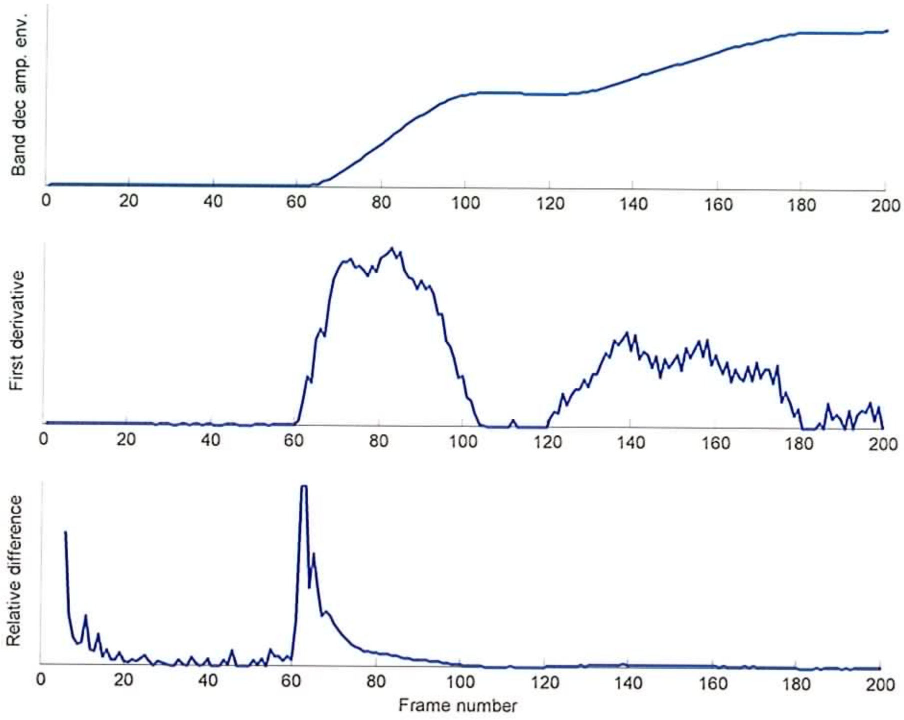

**Figure 4-7: Amplitude envelope (top plot), first derivative (middle plot) and relative first derivative (bottom plot) of a signal**

An extensive literature review of energy based onset detectors has been given in this section. The range of reviewed methods includes standard energy based methods [Chafe '86], high frequency methods [Bilmes '93, Masri '96b], multiband approaches [Klapuri '98, '99, Scheirer '98], and a spectral difference method [Masri '96a]. In Section 4.5, the benefits and disadvantages of using these methods against phase based and combined energy and phase based methods is discussed.

## 4.3 Phase based onset detection

A significant amount of work in onset detectors based on phase vocoder theory has been performed [Dolson '86]. In this section, the basic concepts of phase vocoder theory are first explained, followed by a review of the approaches that have incorporated it into their systems.

### 4.3.1 Phase vocoder theory

The short-time Fourier Transform (STFT) of the discrete signal $x(m)$ is given by:

$$X(n,k) = \sum_{m=-\infty}^{\infty} x(m)h(n-m) * e^{-j(2\pi\ \ N)k.m} = |X(n,k)| * e^{j\varphi(n,k)} \tag{23}$$

where $n$ and $k$ are the hop number and frequency bins respectively

If a stable sinusoid with frequency $\Omega_k$ exists for a given $k$ value, the unwrapped target phase $\widetilde{\varphi}_t(n,k)$ can be calculated using the unwrapped bin phase of the previous hop for that $k$ value, which is denoted as $\widetilde{\varphi}(n-1,k)$:

$$\widetilde{\varphi}_t(n,k) = \widetilde{\varphi}(n-1,k) + \Omega_k h \tag{24}$$

where $h$ is the hop size

However, the above equation only recreates the ideal case of using synthetic signals with frequencies corresponding to the frequency bin. Therefore, a phase deviation $\widetilde{\varphi}_d(n,k)$ is expected:

$$\widetilde{\varphi}_d(n,k) = princarg[\widetilde{\varphi}(n,k) - \widetilde{\varphi}_t(n,k)] \tag{25}$$

where princarg is the principal argument function mapping the phase into the [-π:π] range.

Thus, the unwrapped phase will be given by:

$$\tilde{\varphi}_u(n,k) = \tilde{\varphi}_t(n,k) + \tilde{\varphi}_d(n,k) \tag{26}$$

We can then calculate the phase increment per hop:

$$\Delta\varphi_h(n,k) = \tilde{\varphi}_u(n,k) - \tilde{\varphi}(n-1,k) = \Omega_k h + \tilde{\varphi}_d(n,k) \tag{27}$$

The instantaneous frequency will be then given by:

$$f_i(n,k) = \frac{\Delta\varphi_h(n,k)}{2\pi h} f_s \tag{28}$$

where $f_s$ is the sampling rate.

### 4.3.2 Phase based onset detection approaches

By looking at the instantaneous frequency difference of the frequency bins between consecutive frames $\Delta f_i(n,k)$, and therefore the phase increment of the frequency bins on a frame-by-by frame basis, steady and transient bin components can be separated [Settel '94].

$$|\Delta f_i(n,k)| = f_i(n,k) - f_i(n-1,k) = \frac{(\Delta\varphi_h(n,k) - \Delta\varphi_h(n-1,k))f_s}{2\pi h} \tag{29}$$

In the case of a steady component, the sinusoid is stable and it is expected that the instantaneous frequency, and therefore the phase increment, have very similar values between two adjacent frames.

Thus, $|\Delta fi(n,k)| = 0$ during the steady state

$|\Delta fi(n,k)| > 0$ during a transient state

From Equation (29), the differential angle $d\varphi(n,k)$ can be obtained:

$$d\varphi(n,\kappa) = |\Delta f_i(n,k)| \frac{2\pi h}{f_s} = \Delta\varphi_h(n,k) - \Delta\varphi_h(n-1,k) \tag{30}$$

Then, by combining Equations (30), (27), (25) and (24), $d\varphi(n,k)$ is given by:

$$d\varphi(n,\kappa) = \mathrm{princarg}\left[\widetilde{\varphi}(n,k) - 2\widetilde{\varphi}(n-1,k) + \widetilde{\varphi}(n-2,k)\right] \tag{31}$$

where $d\varphi$ corresponds to a measure of the phase deviation between target and current frame.

Thus, by allowing a threshold $Tss$, the analysed component is steady if:

$$d\varphi < Tss \tag{32}$$

On the other hand, if the component is a transient, the sinusoid is not stable because an unpredicted phase value has occurred:

$$d\varphi > Tt \tag{33}$$

where $Tt$ is the transient detection threshold.

As an example we consider Figure 4-8, where the above method has been applied to separate transient bins (middle plot) from stable bins (bottom plot) taken from a piano signal (top plot). It can be appreciated that the transient bins concentrate more in the onset part of the signal. [Settel '94]´s approach does not perform onset detection. However, the measure of the differential angle (Equation (31)) forms the basis of future onset detection approaches.

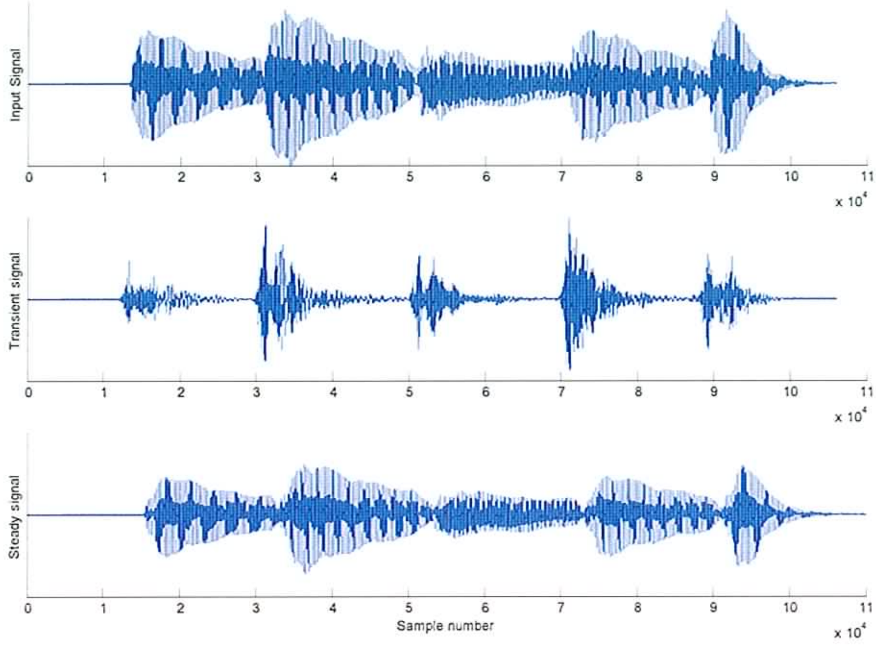

**Figure 4-8: Separation of transient bins (middle plot) and stable bins (bottom plot) from a polyphonic piano tune (top plot)**

In [Duxbury '01], another system to separate transients from steady bins based on [Settel '94]'s approach is presented. The method utilises a constant Q filter bank implementation in order to split the signal into six octaves. As in [Settel '94], the previously explained phase vocoder theory is then applied to separate the transient bins in each band. The analysis is implemented using a band frequency dependent window length and an adaptive threshold $A_{SS,}$ which takes into consideration previous frames. Thus, if a bin $k$ has been steady in the previous frames, and its $d\varphi$ value is just above the original $T_{SS}$ value, the threshold is increased further to allow the bin to remain steady:

$$A_{ss}(n,k) = T_{ss} + \alpha(n,k)T_{ss} + \beta(n,k)T_{ss} \qquad (34)$$

where $\alpha(n,k)$ and $\beta(n,k)$ are given by:

$$\alpha(n,k)=\begin{cases}0 \text{ for } & \Delta\varphi(n-1,k)-\Delta\varphi(n-2,k)<A_{ss}(n-1,k)\\ a \text{ for } & \Delta\varphi(n-1,k)-\Delta\varphi(n-2,k)>A_{ss}(n-1,k)\end{cases} \qquad (35)$$

and

$$\beta(n,k)=\begin{cases}0 & \Delta\varphi(n-2,k)-\Delta\varphi(n-3,k)<A_{ss}(n-2,k)\\ b & \Delta\varphi(n-2,k)-\Delta\varphi(n-3,k)>A_{ss}(n-2,k)\end{cases} \qquad (36)$$

where $T_{ss}$ is a threshold set up by the user, and $a$ and $b$ are real numbers

This approach does not provide the onset times. However, it reduces the amount of energy in the steady state, since the onset information is concentrated in the transient bins [Duxbury '01]. Thus, in order to improve the accuracy of the onset detection, Duxbury et al. suggest combining the transient separation with standard energy based onset detection methods [Duxbury '01]. However, this is only the case for a sharp onset, since a slow onset does not have such fast change in the onset part. The system introduces the principle of using a frequency band threshold. Nevertheless, it still requires a user input to enter the fixed threshold $T_{ss}$.

[Bello '03] also based his approach on phase vocoder theory to generate a frame-by-frame statistical distribution of differential angles for all $k$ (Equation (31)). The distributions are expected to vary in the different stages of the signal. In the absence of transients, the histogram has a normal distribution. When the onset occurs, the difference between the target and actual phase of the frame bins increases (Equation (29)), spreading the distribution. During the steady part of the signal, the sharpness and height of the distribution increases.

This is illustrated in Figure 4-9, where a histogram of a 512 samples frame of the onset part (samples 9729 to 10240), and a histogram of a 512 samples frame of the harmonic part (samples 29185 to 29696) of the piano signal depicted in the top plot, are depicted in the middle plot and bottom plot respectively.

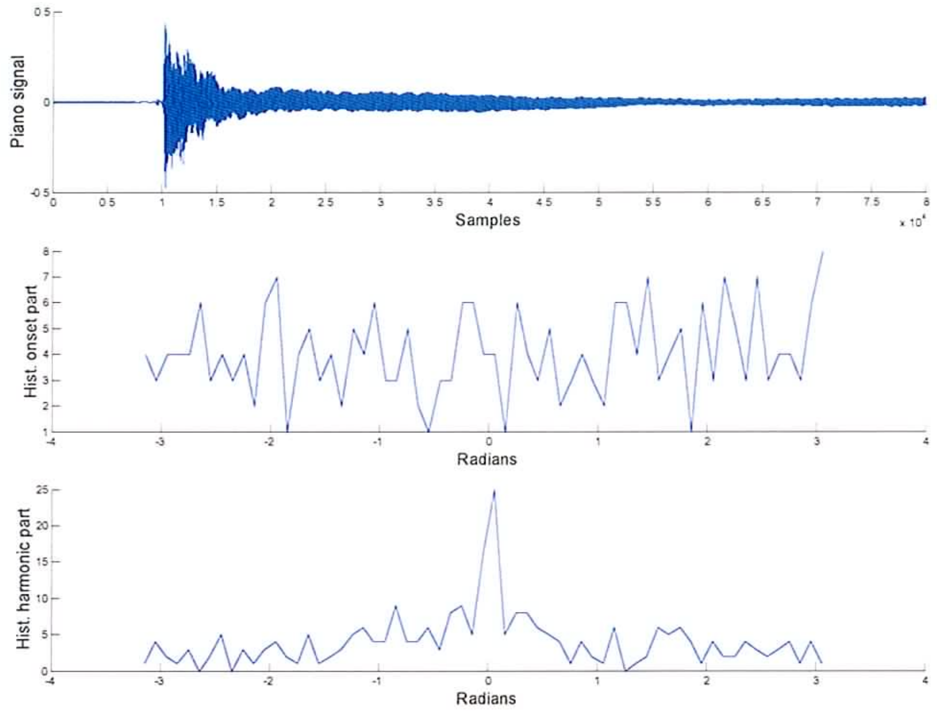

**Figure 4-9 Histogram of the phase deviation distribution in the onset part (middle plot), harmonic part (low plot) of a piano signal (top plot)**

To measure the spread of the distribution per frame, Bello uses the Inter Quartile Range (IQR) method [Pal '05], which computes the difference between the 75th and the 25th percentiles of the data being analysed. As can be appreciated in Figure 4-10, where the distribution of the phase spread of a piano signal is depicted, the IQR (C plot) gives a slightly smoother representation than the standard deviation (B plot).

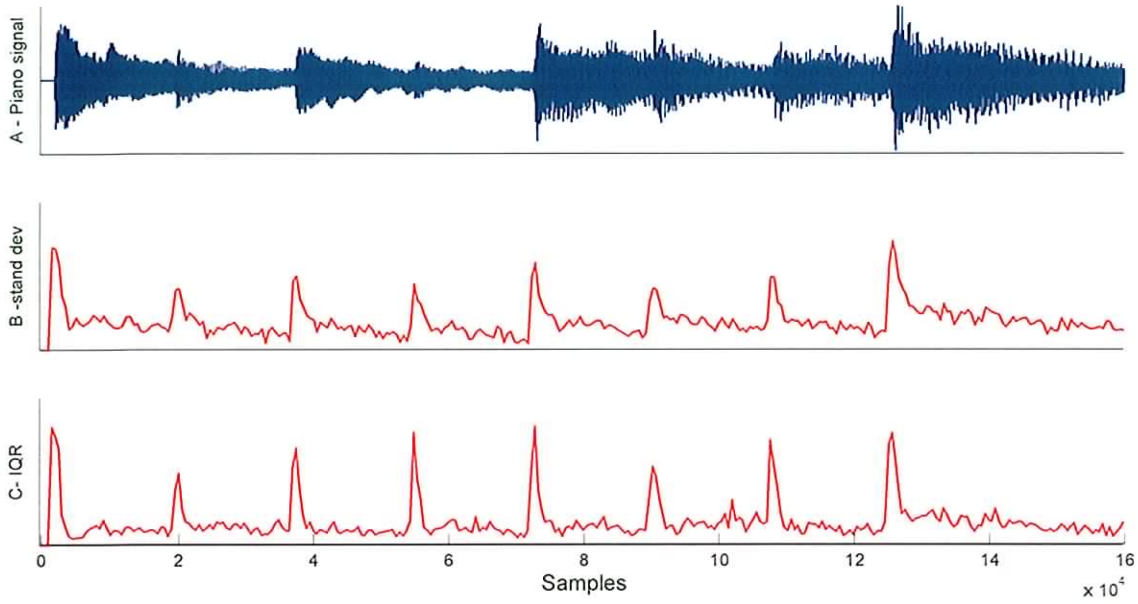

**Figure 4-10: Standard deviation (B top) and IQR (C plot) phase based onset detection function of a piano excerpt of "Hold on" by Shinya Iguchi (A plot)**

Bello identifies the beginning of the steady state of the signal using the Kurtosis method, which is a measure of how sharp a distribution is [Bulmer '79].

$$\gamma_2 = \frac{\mu_{4(\mu)}}{\sigma^4} - 3 \qquad (37)$$

where $\mu_4$ and $\sigma$ are the fourth central moment and the standard deviation respectively. Thus, the detected onset will be at the closest peak (onset candidate) in the IQR function to the Kurtosis function (steady state).

A dynamic threshold $\delta_t(m)$ is calculated based on the weighted median average of the Kurtosis detection function $\gamma 2$, within a length $H$ sliding analysis window $K_m$, given by:

$$\delta_t(m) = C_t \, median\gamma 2(K_m), K_m \in [m - \frac{H}{2}, m + \frac{H}{2}] \qquad (38)$$

In [Bello '04], Equation (38) is replaced by:

$$\delta_t(m) = \delta + C_t median \gamma 2(K_m)$$

(39)

where $C_t = 1$ and $\delta$ is varied in small steps.

In both equations, the onset detection results for different weighting values are compared against a database of hand labelled onsets. Then, the set of parameters that obtains the best detection results is chosen [Bello '03, Bello '04].

In this section, methods that utilise phase vocoder theory to deal with the onset part of the signal have been introduced. First, a brief description of phase vocoder theory has been given. Then, two methods to separate transients from steady bins using phase vocoder theory have been explained, which is useful when isolating sharp onsets [Settel '94]. Finally, an onset detection method based on the statistical distribution of the differential angles obtained by using phase vocoder is also reviewed [Bello '03]. In Section 4.5, the benefits and disadvantages of using these methods against phase based and combined energy and phase based methods is discussed.

## 4.4 Combining phase and energy approaches

[Duxbury '02] proposes a hybrid approach that uses different methods in high and low subbands for detecting different types of onsets. The signal is first split into four bands. The highest two bands (2.5-11 kHz) use energy based methods, which are useful for detecting "hard onsets".

The highest band, $S_1$, uses standard energy detection in order to detect the energy of the frame (SE-Equation (15)). In order to avoid multiple detections in noisy regions located after the onset, the band detection function considers previous frames, but giving more weight to the most recent values as follows:

$$ons(n) = SE(n) - \sum_{a=1}^{A} \frac{SE(n-1)}{a} \tag{40}$$

In upper middle frequency subband $S_2$, a method previously suggested in [Duxbury '01] is applied. First, the band transient bins $K_{tr}$ are separated from the steady bins by using phase vocoder theory (Equation (33)). Then, the energy of the transient bins is calculated as follows:

$$TE(n) = \sum_{k \in K_{tr}} | X(k, nh) |^2 \tag{41}$$

In the two lowest subbands $S_3$ and $S_4$, a function based on the spectral difference method of [Masri '96a] (Equations 18) and 19)) is applied and is given by:

$$DM(n) = \sum_{k=1}^{N/2} dX_n(k)^2 \tag{42}$$

where $dX_n(k)$ is given by Equation (18).

As a comparison, Equation (42) is applied to the same Banjo signal as in Figure 4-4. The detection function is depicted in Figure 4-11, where it can be appreciated that the detection function is slightly smoother than Masri's [Masri '96a] (Figure 4-4).

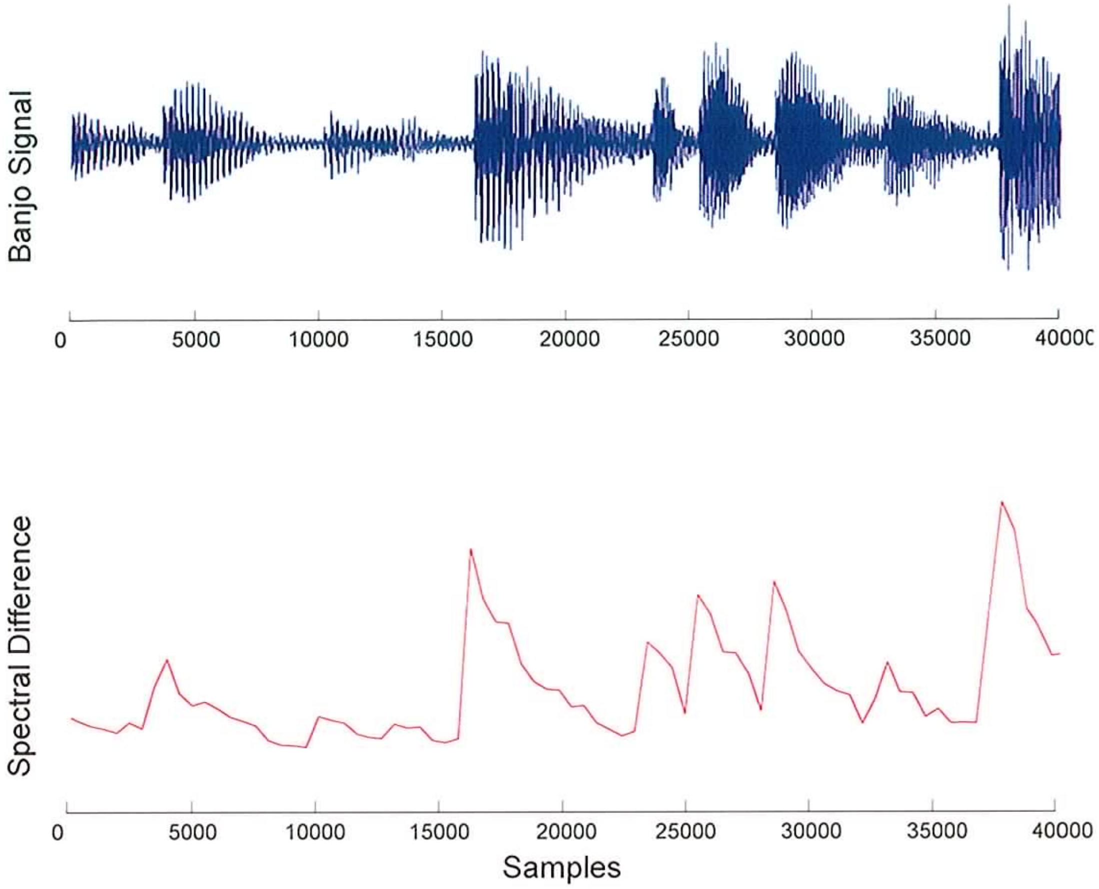

**Figure 4-11: Duxbury's spectral difference onset detection function of a banjo excerpt of "The Cock and the Hen", played by RIRA**

In order to detect softer onsets in a similar manner to Equation (22), the detected measure is normalised as follows:

$$DM = \frac{\sum_{\{k;dX_n(k)>0\}} dX_n(k)^2}{\sum_{k=1}^{N/2} | X(k,(n-1)h) |^2} \qquad (43)$$

where the distance measure has only been considered for the positive values in order to reject offset detection.

This hybrid method uses a statistical approach to set the threshold. The histogram of a 5s sliding window of the detection function is calculated, which could be seen as a combination of two probability functions: the probability of having a transient, $p(tr)$, which are defined as outliers in the histogram, and the probability of not having a transient, $p(nt)$, which has a narrow distribution. Thus, if a transient occurs, it is expected that the standard deviation is larger than when a transient is not present. The idea behind this method is to find the point in the histogram where the data is more likely to be an onset, which they experimentally set at the maximum of the second derivative of the histogram.

Next, the detected onsets in every band are combined, giving more weight to the higher frequency bands since they provide better time resolution.

$$P(t) = \alpha P_{S1}(t) + \beta P_{S2}(t) + \gamma P_{S3}(t) + P_{S4}(t) \tag{44}$$

where α>β>γ>1 and  is a time vector containing ones (1's) at onset positions and zeros (0's) elsewhere for the band *i*.

As in [Klapuri '98], the most prominent onset within a 50 ms window is kept. An advantage of this method is that by changing the weighting coefficient values, the system can be tuned to detect sharp or slow onsets [Duxbury '02].

In [Duxbury '03a, b], Duxbury et al. combine phase and amplitude deviation in the same representation to exploit their ability to detect soft and weak onsets respectively.

In [Duxbury '03a], to quantify the energy and phase frame spread, the probability density function $f_n(x)$ of the energy deviation vector (Equation (18)), and the phase deviation vector are calculated independently per frame (Equation (31)). In order to measure the distribution, the mean replaces the Kurtosis calculation of Equation (37):

$$\eta(n) = mean(f_n(|x|)) \tag{45}$$

Finally, the calculations of the spread of the energy distribution $\eta_e(n)$ and the phase distribution $\eta_p(n)$ are combined as follows:

$$\eta_t(n) = \eta_e(n) . \eta_p(n) \tag{46}$$

and is compared to a threshold value as in Equation 22.

As an example, by using Equation (46) the onset detection function of the same piano signal as in Figure 4-10 is depicted in Figure 4-12, plot B. It can be appreciated that the function is less noisy than any of the phase based onset detection functions of Figure 4-10. By contrast, an energy based onset detection function obtained by utilising Equation (21) is depicted, where it can be appreciated that many spurious detections arise.

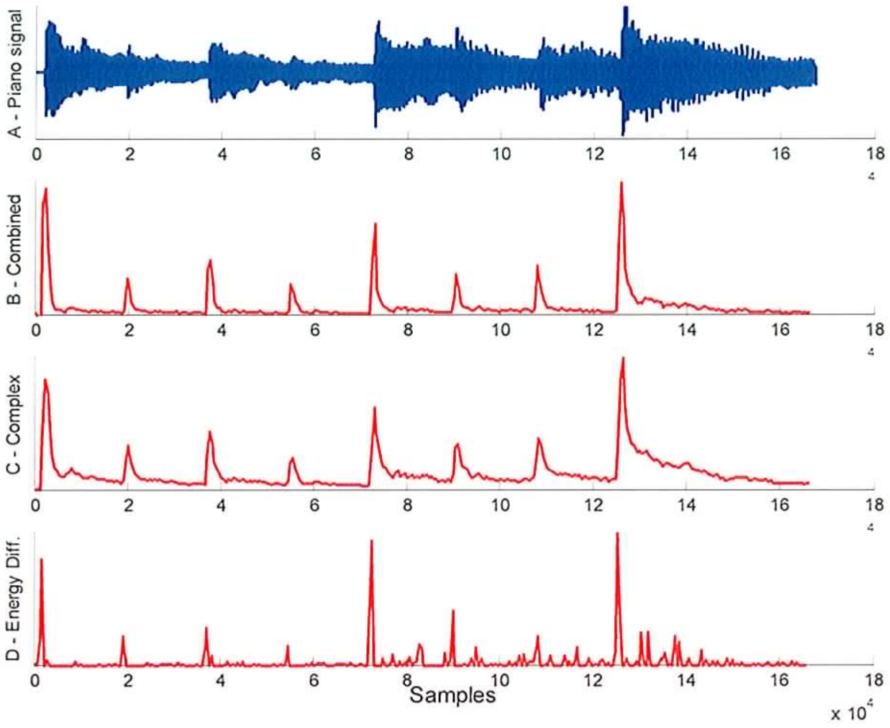

**Figure 4-12: Combined (B plot), complex based (C plot) and energy based detection function (D plot) of a piano excerpt of "Hold on" by Shinya Iguchi (A plot)**

[Duxbury '03b] combines phase and amplitude information to develop a complex domain approach.

The target value for a FFT $k_{th}$ bin is:

$$\widehat{S}_k(m) = \widehat{R}_k(m)e^{j\hat{\phi}_k(m)} \qquad (47)$$

where for the steady state of the signal:

$\widehat{R}_k(m)$ is the bin magnitude of the frame, which is expected to be the same as the magnitude of the previous frame for the same $k$ bin, $R_k(m\text{-}1)$.

$\hat{\phi}_k(m)$ is the expected phase for the $k_{th}$ bin, which is equal to $2*\widetilde{\varphi}(n-1,k)-\widetilde{\varphi}(n-2,k)$

From Equation (31), the above expected phase value produces a differential angle equal to zero.

The measured value for the same $k_{th}$ bin is:

$$S_k(m) = R_k(m)e^{j\phi_k(m)} \tag{48}$$

As can be appreciated in Figure 4-13, by using the Euclidean distance the deviation from the target and measured $k_{th}$ bin can be estimated as:

$$\Gamma_k(m) = \left\{[\Re(\hat{S}_k(m)) - \Re(S_k(m))]^2 + \ldots + [\Im(\hat{S}_k(m)) - \Im(S_k(m))]^2\right\}^{1/2} \tag{49}$$

where $\Re$ and $\Im$ are the real and imaginary parts respectively.

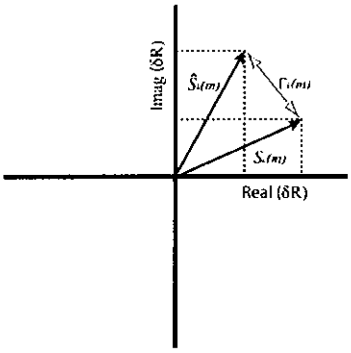

**Figure 4-13: Target and current complex domain vectors, and the Euclidean distance between them**

Then, summing all the measures across all $k$, the frame-by-frame deviation can also be determined:

$$\eta(m) = \sum_{k=1}^{K} \Gamma_k(m) \tag{50}$$

The same threshold function as that in [Bello '03] was used.

As an example, the onset detection function of the same piano signal as in Figure 4-10 is obtained by using the complex detection method, and is depicted in the Figure 4-12, plot C.

In order to combine the time resolution of higher frequencies, and the noise robustness of the lower ones, a multi resolution version of [Duxbury '03b] is suggested in [Duxbury '04]. As in [Duxbury '02], the method utilises four bands, in which the complex based method is applied separately in each band. Since at high frequencies percussive instruments produce more prominent energy changes (Figure 4-1), high bands provide an accurate estimation of the time onset localisation by using short frame lenghts [Duxbury '04]. By contrast, low frequency bands are more adequate to detect signal changes. However, this requires the use of long frame lengths in slow onset signals, which has the cost of poorer time resolution [Duxbury '04].

In this section, methods that combine information of the phase and energy of the signal to estimate onset times have been reviewed. In [Duxbury '02], the energy of the transient bins is calculated, which have been previously separated from the steady bins utilising phase vocoder theory. In [Duxbury '03a, b], phase and amplitude deviation are compounded to form a unique deviation value per bin. In [Duxbury '03a], the statistical distribution of the differential angles and amplitude are multiplied. In [Duxbury '03b], the predicted amplitude and phase is combined to form a complex number. [Duxbury '02, Duxbury '04] incorporate a multi band decomposition in the analysis. In Section 4.5, the benefits and disadvantages of using these methods against phase based and energy based methods is discussed

## 4.5 Discussion

The significant amount of research that has been undertaken on onset detection has been described in Sections 4.2, 4.3 and 4.4. Existing approaches generally perform successfully on detecting sharp onsets. However, their performance considerably degrades if the onset has a different profile. In addition, energy based methods are prone to spurious onset detections when dealing with notes modulated in amplitude. Energy methods that give more emphasis to high frequency bins perform well for sharp onsets. However, as shown in Figure 4-3, the accuracy is highly related to the characteristics of the signal [Bilmes '93, Masri '96b]. A problem related to methods that use the first order difference to obtain the onset detection function, as in [Scheirer '98], is that the maximum in the function is delayed from the actual onset time. The relative function of [Klapuri '98] provides a better onset time estimation. However, as shown in Figure 4-7, spurious onset detections can arise as a result of applying this function, since a small amplitude increase between small amplitude envelope values can be excessively high in relation to the amplitude envelope.

Energy based approaches encounter problems in fast transitions between notes that do not produce a significant energy increase. This is partially solved by using a multiband approach. Nevertheless, the number of bands utilised in those methods is not large enough for resolving fast transitions between very close notes: [Klapuri '99, Scheirer '98] use octave filter bands and [Klapuri '99] utilised third octave filter bands. In addition, [Klapuri '99] calculates the intensity of the onset candidates by multiplying the onset prominence value by the band centre frequency. This gives more weight to high frequency bands, thus favouring percussive onsets as opposed to slow onsets. In addition, [Klapuri '98, '99] combine the onset candidates in all frequency bands by summing their prominence values within a 50 ms time window. This is not appropriate for slow onsets that have energy in only a few harmonics, because it would only produce peaks in a small number of bands.

Phase based methods give more prominent detection peaks than energy based methods in signals that have weak high energy [Bello '04]. However, the method is significantly sensitive to noise, which increases the number of spurious detections. When the magnitudes contain a very small value, their corresponding phases take random values [Smith '97], which will produce incoherent results by performing phase based onset detection. In order to avoid this problem, a small amplitude threshold is also required, and it has been utilised in all the examples illustrated in this section. Another limitation of the phase based approaches is on dealing with signals with frequency modulations. When analysing the steady state of the signal, they assume that the instantaneous frequencies between two adjacent frames have very similar values (Equation (29)), which is not the case during frequency modulations.

Combined energy and phase approaches merge the benefits and disadvantages of both approaches. They provide better estimations when dealing with slow onsets than energy based approaches, and produce smoother onset detection functions than phase based approaches. However, the methods are still prone to detecting spurious onsets when amplitude and frequency modulations are present in the signal.

Multi-band approaches have also been used in methods that incorporate phase information [Duxbury '02, Duxbury '04]. Both methods divide the frequency range into only four bands. At high frequencies, percussive onsets produce a sharp increase of energy. Thus, by using short frame lengths in the analysis, high bands will provide good time resolution. However, since slow onset signals do not provide such energy burst, high frequency bands will not improve the slow onset detection accuracy.

A good advantage of systems based on the spectral difference [Masri '96a], and of phase vocoder based approaches as [Duxbury '03a, b], is that they analyse the signal bin per bin, thus being sensitive to all bin signal changes.

Generally, peaks in a detection function that reach a given threshold are considered onset candidates. In [Bilmes '93, Chafe '86, Klapuri '98, Klapuri '03 , Masri '96a, Masri '96b], the threshold is set manually by the user. In [Bello '03, Duxbury '02], two different approaches to set the threshold automatically are provided.

## 4.6 Conclusions

In this chapter, the main existing onset detection methods have been reviewed, which includes energy, phase, and combined energy and phase based approaches. The advantages and disadvantages of using the methods have been discussed in Section 4.5. It is apparent that a robust system capable of dealing with both amplitude and frequency modulation is yet to be implemented. In addition, accurate onset detection of slow onsets also remains an open issue.

As introduced in Chapter 2, Irish traditional music is commonly played with inclusion of ornamentation. Existing approaches will encounter problems detecting ornamentations such as cuts or strikes. In these cases, the ornamentation and the onset events can occur separated by a very short space of time, and both events can be estimated as a unique candidate. Problems also arise in analysing fast passages such as legatos, which are very common in Irish traditional music. In this case, the increase of energy of the new event can be very small, which can cause problems when using energy based onset detectors. In addition, as mentioned in the introduction, the majority of Irish traditional music instruments such as the tin whistle, fiddle, flute, concertina and uilleann pipe have a slow onset.

In order to overcome the identified problems on detecting slow onsets modulated in amplitude and frequency, two different approaches are presented. First, a method which focuses on the characteristics of the tin whistle within Irish traditional music is presented in Chapter 6 [Gainza '04c], which represents Contribution 1 in Section 1.2.1. The problems related to legato playing are reduced by using an energy based multi-band approach, which uses one specific band per note. This is also an advantage for the purpose of detecting ornamentation events, since they will arise in a separate band to the note they ornament. The Irish tin whistle can produce strong amplitude modulations, which can cause problems when using energy based onset detectors, since amplitude modulations in high notes can have energy increases with a similar value to onset energy increases in low notes. In addition, each note in a slow instrument is played at a different pressure range, which increases with frequency [Martin '94]. Thus, in order to reduce the effect that amplitude modulations produce, the use of different automatic band thresholds is investigated. The first method sets the band threshold according to [Martin '94]'s theory, which justifies the use of an energy based approach. The second utilises the standard deviation to pick the onset candidates. Both thresholding methods form the part of Contribution 1.

Existing onset detectors utilise energy and/or phase to generate an onset detection function. A novel onset detector that tracks the harmonicity changes of the signal by using comb filters is presented in Chapter 8 [Gainza '05b], which has been introduced in Chapter 1, Contribution 3. The method is robust in dealing with frequency and/or amplitude modulations. In addition, the method relates the harmonicity detection to the energy of the analysing frame, which is suitable to detect slow onsets.

In the following chapter, a literature review of existing pitch detection methods is provided. Pitch detection in conjunction with onset detection forms the core of music transcription algorithms.

# 5 Pitch detection

## 5.1 Introduction

In the previous chapter, existing onset detection techniques have been introduced. A discussion of the methods has also been provided which lead to the development of two novel onset detection methods. This chapter focuses on pitch detection, which plays a crucial part in music transcription, aiming to estimate the notes that comprise a given audio segment. It is a very common task in music transcription systems to first segment the audio signal by using an onset detector. Then, the pitches that compose the audio segment are estimated by using a separate pitch detection block. This shows that onset and pitch detection are interrelated disciplines within music transcription

There are numerous methods that perform pitch detection. In order to illustrate such diversity, heterogeneous models based on detecting the periodicity of the time and frequency domain, auditory modelling, knowledge modelling or data representations are introduced in this chapter. The difficulty of classifying the methods should be noted, where some of them could fall into more than one of the above mentioned categories. At the end of the chapter, a discussion of the methods and conclusions will also be given.

## 5.2 Time domain periodicity methods

In this section, a review of time domain periodicity based methods is given. First, comb filters/AMDF techniques for the use of pitch detection are introduced in Section 5.2.1. A general explanation of comb filter techniques was previously introduced in Chapter 3. These techniques are of particular interest in this thesis, since they form the basis of the work implemented in onset detection (Chapter 8) and pitch detection (Chapter 9). Further, other time domain periodicity methods are introduced in Section 5.2.2. Finally, the frequency domain characteristics of time domain periodicity methods, and in particular of comb filter techniques, are explained in Section 5.2.3. In addition, methods to sharpen the magnitude response of the comb filters are also introduced.

### 5.2.1 Comb Filters /AMDF techniques

Moorer pioneered the use of comb filters for the purpose of speech pitch detection [Moorer '74]. This approach determines the FIR comb filter with $b_l$ = -1 that when applied to the signal produces a minimum in the output energy $y(D)$. By applying the method within a frame of $N$ samples, $y(D)$ is given by:

$$y(D) = \sum_{n=D+1}^{N} [x(n) - x(n-D)]^2 \qquad (51)$$

Moorer calculates $y(D)$ for the range of delays which correspond to the pitch range of the human voice (70 – 225 Hz), which is the delay $D$ range 196 - 630 samples for the case of a sampling frequency equal to 44100 Hz. As an example, the same method is applied to a synthetic E4 tone (Figure 5-1, top plot), which has five harmonics with amplitude ratio equal to the fundamental harmonic amplitude divided by the harmonic number. The resulting output is shown in the bottom plot of Figure 5-1, where it can be appreciated that there is a minimum in the delay sample $D$ = 134, which corresponds to the E4 pitch period in samples. Other minima also appear at integer multiples of the pitch period, which can cause detection errors in lower octave multiples of the pitch. Thus, only the smallest delay minimum is used to determine the pitch. To improve the speed efficiency of the system, only filters that provide a frequency resolution of 1 Hz are utilised. For example, delays 621 and 630 correspond to frequencies 71.01 and 70 Hz respectively. Thus, filters within the 622 -629 delay range are not used (as 71.01 – 70 ≈ 1 Hz).

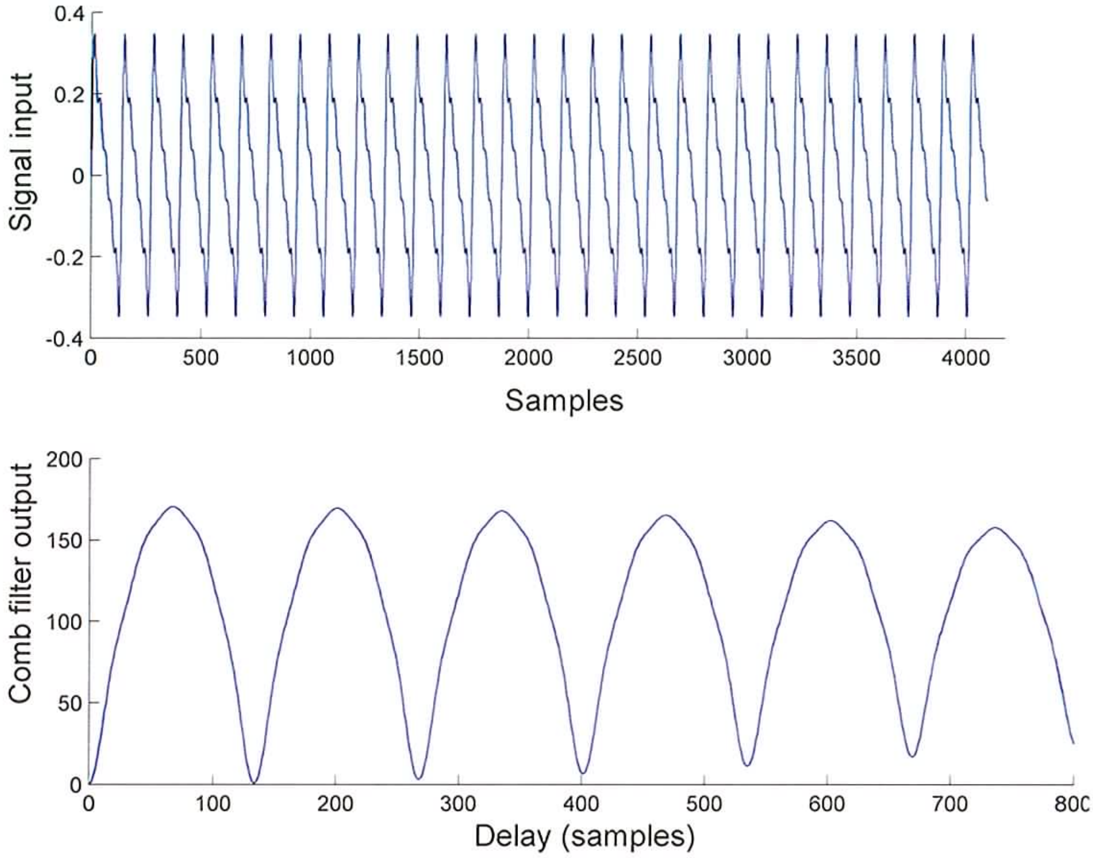

**Figure 5-1: Energy comb filter output of an E4 tone with 5 harmonics**

In [Moorer '75], a system that utilises FIR comb filters to detect musical chords is presented. The system investigates the position of the minima in $y(D)$ after applying the comb filters. An example is illustrated in Figure 5-2, where the C4 major triad, which is composed of the notes C4, E4 and G4, is analysed using the above method. Local minima arise at filter delays equal to 166, 134 and 111 samples, which correspond to notes C4, E4 and G4 respectively. However, it can be appreciated that the strongest minimum appears at a filter delay equal to 673, which corresponds to C2. This note has its 4th, 5th and 6th harmonics located at the frequencies of C4, E4 and G4 respectively.

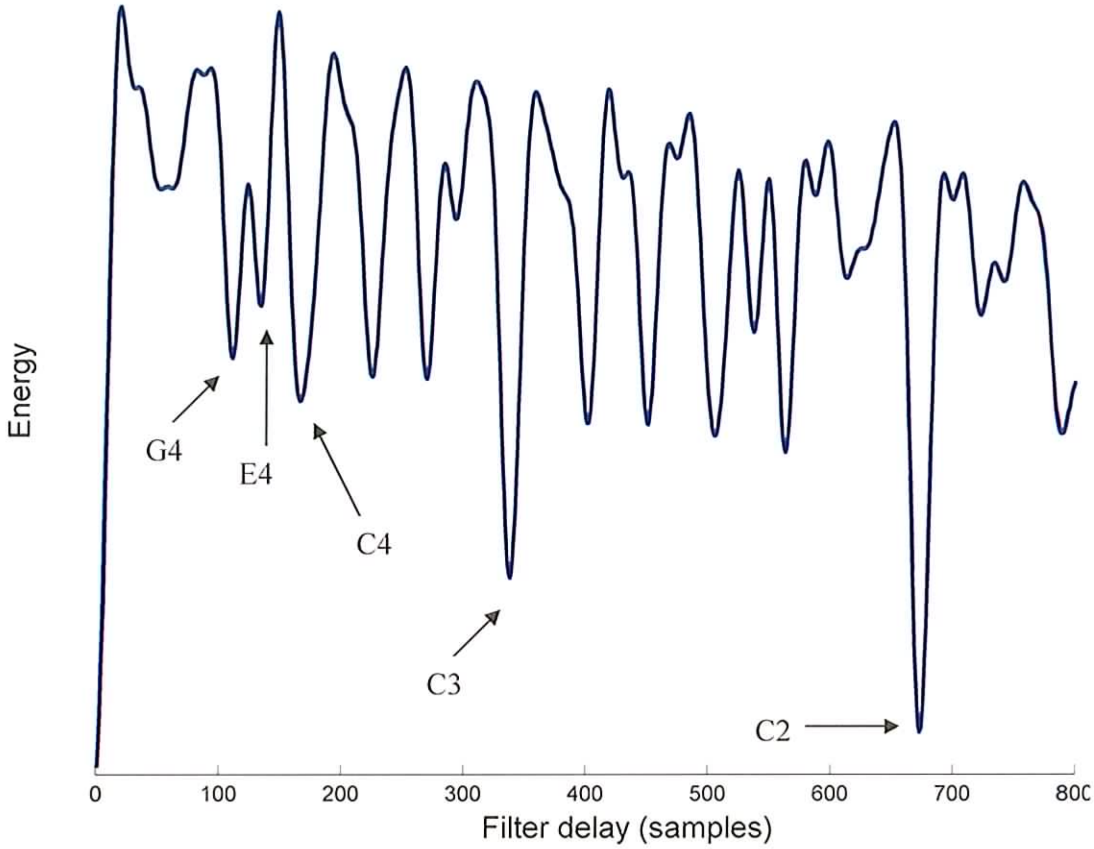

**Figure 5-2: Example of the use of Moorer's chord detector**

Another approach named Average Magnitude Difference Function (AMDF) measured the pitch signal by summing the difference between the input signal $x$ and a delayed version of the same signal using different delays $D$. The delay at which a minimum occurs, corresponds to the pitch period [De Cheveigne '02, Ross '74].

$$AMDF(D) = \frac{1}{N}\sum_{n=D+1}^{N}\left|x(n) - x(n-D)\right| \tag{52}$$

where $N$ is the frame length.

As can be appreciated from Equation (3), the AMDF approach is equivalent to a comb filter with $b_l$ = -1. The same method is utilised by [De Cheveigne '91b] to detect the pitch of two speech signals by connecting two AMDF comb filters in cascade using different filter delays. Thus, the delay combination that produces a minimum in the output provides the period of the two speech signals.

In [Miwa '99a; Miwa '00], FIR comb filters are utilised for the purpose of music transcription within octaves 3 to 5, where octave 4 is the octave beginning at middle C. Miwa built 12 comb filters with $b_l$ = -1, one for each note of the octave 3, where the first notch of each FIR comb filter exactly matches the frequency of a different note of octave 3, which are a semitone apart on a tempered scale. The filters are then connected in cascade, and the filter which makes a zero output represents the detected note. This can be appreciated in Figure 5-3, where $H(i,j)$ is the transfer function of a note $j$ in octave $i$. When a new note is detected, the existence of other notes in the audio signal is investigated. This is performed by moving the filter that represents the detected note to the first position in the cascade configuration, and then applying the cascade filtering operation again. This procedure is performed iteratively until the $y12(n)$ output has a non zero value, which signifies that all the notes have been detected.

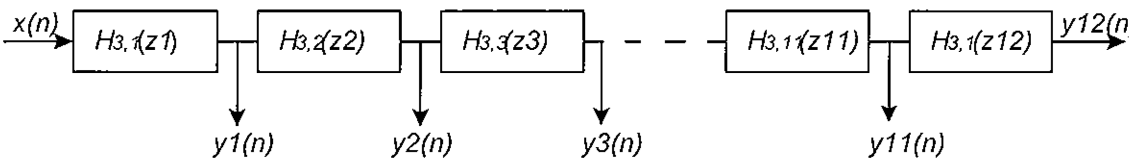

Figure 5-3: Cascade connection of FIR comb filters.

In [Tadokoro '02], a method that improves the performance of [Miwa '99a, Miwa '00] in pitch deviations is presented. He utilises three adaptive comb filters in cascade in a musical mix composed of a maximum of three harmonic sounds, as illustrated in the block diagram of Figure 5-4. The delay $D_i$ of the filter $i$ is obtained by calculating the distance between the two maximum consecutive signal peaks of a 10 ms frame. First, the filter delays are initially configured by applying the above method in three consecutive frames of the signal input. Then, the input signal goes through the three pre-configured filters, and the output is utilised to obtain the delay $D_1$ by applying the same method. Finally, the delays $D_2$ and $D_3$ are also calculated from the output, which has been altered by $H_1(z)$ and $H_1(z)*H_2(z)$ respectively. The estimation of $D_1$. $D_2$ and $D_3$ is repeated iteratively until the output energy is below a given threshold.

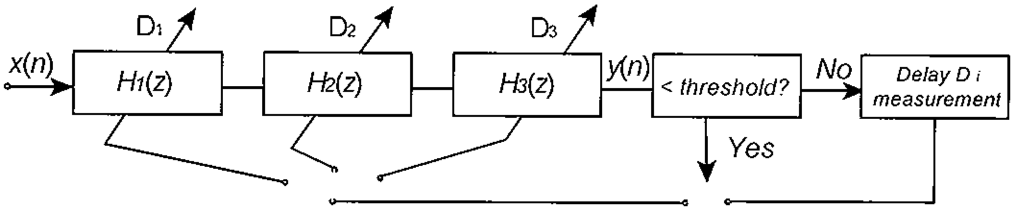

**Figure 5-4: FIR adaptive comb filter algorithm by [Tadokoro '02]**

In order to avoid the signal amplitude alteration caused by applying FIR comb filters in cascade, [Tadokoro '03] proposes another system based on a bank of parallel FIR comb filters, in which the filter that produces an amplitude minimum represents the first detected note. Next, the existence of other notes in the audio signal is investigated by iteratively connecting the output of the filter that has produced the minimum with the input of the parallel comb filter system, and the same filtering process is repeated again until all the notes have been extracted (See Figure 5-5).

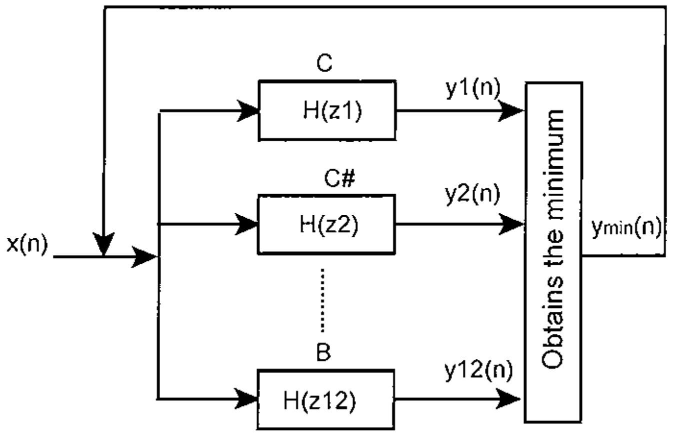

**Figure 5-5: Parallel FIR comb filter configuration [Tadokoro '03]**

Even though [Miwa '99a, Miwa '00, Tadokoro '02, Tadokoro '03] build FIR comb filters for octave 3 notes, their system is designed to detect the pitch of notes within the 3 to 5 octaves range as follows; since two notes one octave apart and two octaves apart are in a 2:1 and 4:1 frequency relationship respectively, the notches of a filter built for a given note *x* in octave 3, will also produce a zero output for the same note *x* in a higher octave. Thus, octave 3 filters will also detect notes played in octaves 4 and 5. This is illustrated in Figure 5-6, where the magnitude response of three standard notch comb filters for detecting the same note in octave 3 (dotted line), octave 4 (dashed line) and octave 5 (solid line) are respectively depicted. If a note *x* is played in octave 3, 4 or 5, with harmonics located at frequency multiples of $f_{3,x}$ $f_{4,x}$ and $f_{5,x}$ respectively, the output will be zero for the three cases if an octave 3 filter is used [Miwa '99a]. Thus, in order to detect the correct octave, [Miwa '99a] suggests the following solution: first, the note played is detected using only octave 3 filters. Then, as illustrated in Figure 5-6, if a note is played in octave 5, the three filters will produce a zero output. If a note is played in octave 4, only the octave 3 and 4 filters will produce a zero output. Finally, if a note is played in octave 3, the output will be zero only if an octave 3 filter was used [Miwa '99a]. This octave detector is also utilised in [Miwa '00, Tadokoro '02, Tadokoro '03] methods.

This method will be successful in the case of a monophonic context. However, in the presence of other co-occurring notes the method will not be as consistent. Since in lower octaves the number of comb filter notches is higher, co-occurring harmonics can be cancelled, thus contributing to produce minimums in low octaves of the filter energy output.

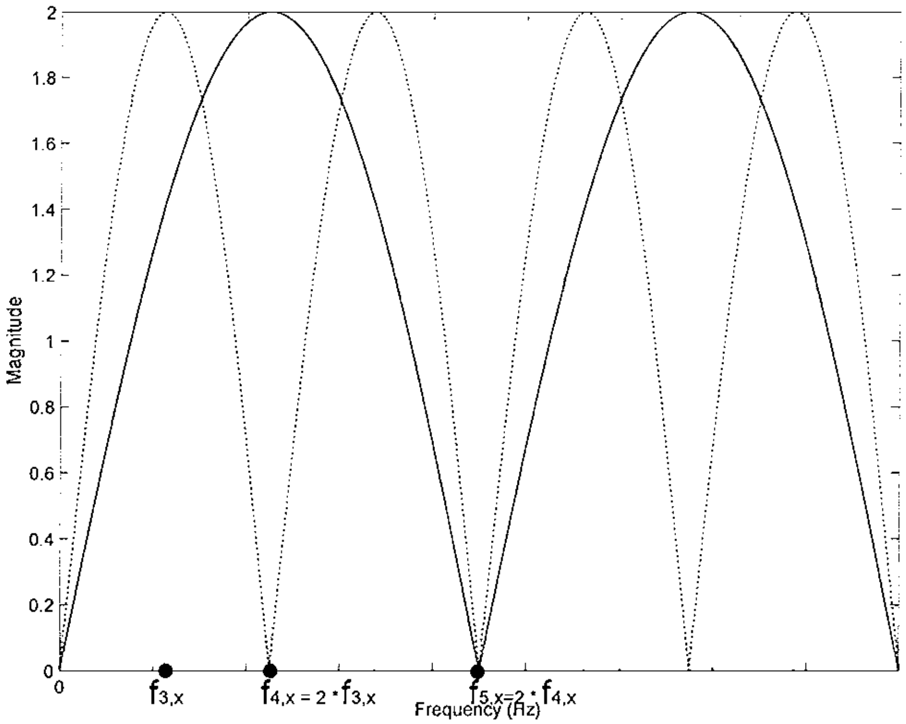

**Figure 5-6: Octave detection comb filter method of [Miwa '99a]**

### 5.2.2 Other time domain periodicity based methods

The autocorrelation method is one of the most widely utilised fundamental frequency estimators [Brown '93]. The autocorrelation function of a frame $N$ is calculated as follows:

$$r_x(D) = \sum_{n=1}^{N-D} x(n)x(n+D) \qquad (53)$$

To ensure that all pitch estimations are reliable, the length of the frame should be bigger than twice the maximum $D$, which also applies to the methods introduced in Section 5.2.1. The resulting function will have peaks at integer multiples of the signal period in the same manner as the comb filter approach.

The autocorrelation function of Equation (53) can be efficiently calculated in the frequency domain using the fast convolution approach as follows [Proakis '95]:

$$r_X = IFFT(|FFT(x(n))|^2) \qquad (54)$$

Cepstrum and autocorrelation methods have certain similarities. By replacing the power spectrum in the autocorrelation function by a logarithm function of the magnitude spectrum, the cepstrum function $c_n$ is obtained as [Martin '82, Noll '64]:

$$c_n = IFFT(\log|FFT(x(n))|) \qquad (55)$$

Thus, Equation (54) emphasises the spectral peak in the presence of noise, and Equation (55) gives more weight to high frequency components. To find a compromise, Tolonen et al utilise the exponent value 0.67 of the magnitude spectrum instead of the power spectrum [Tolonen '00].

Considering Equation (3) with $b_l = 1$ and Equation (53), and ignoring the signal weights, it can be appreciated that a comb filter can be considered as a type of autocorrelation, where summations or subtractions are utilised instead of multiplications, thereby decreasing the computational requirements. This is illustrated in Figure 5-7, where the pitch detection function of the E4 signal (Figure 5-1, top plot) is depicted by utilising the autocorrelation function (top plot) and the comb filter method with $b_l = 1$ (middle plot). For comparison, the cepstrum pitch detection of the same signal is also depicted in the bottom plot.

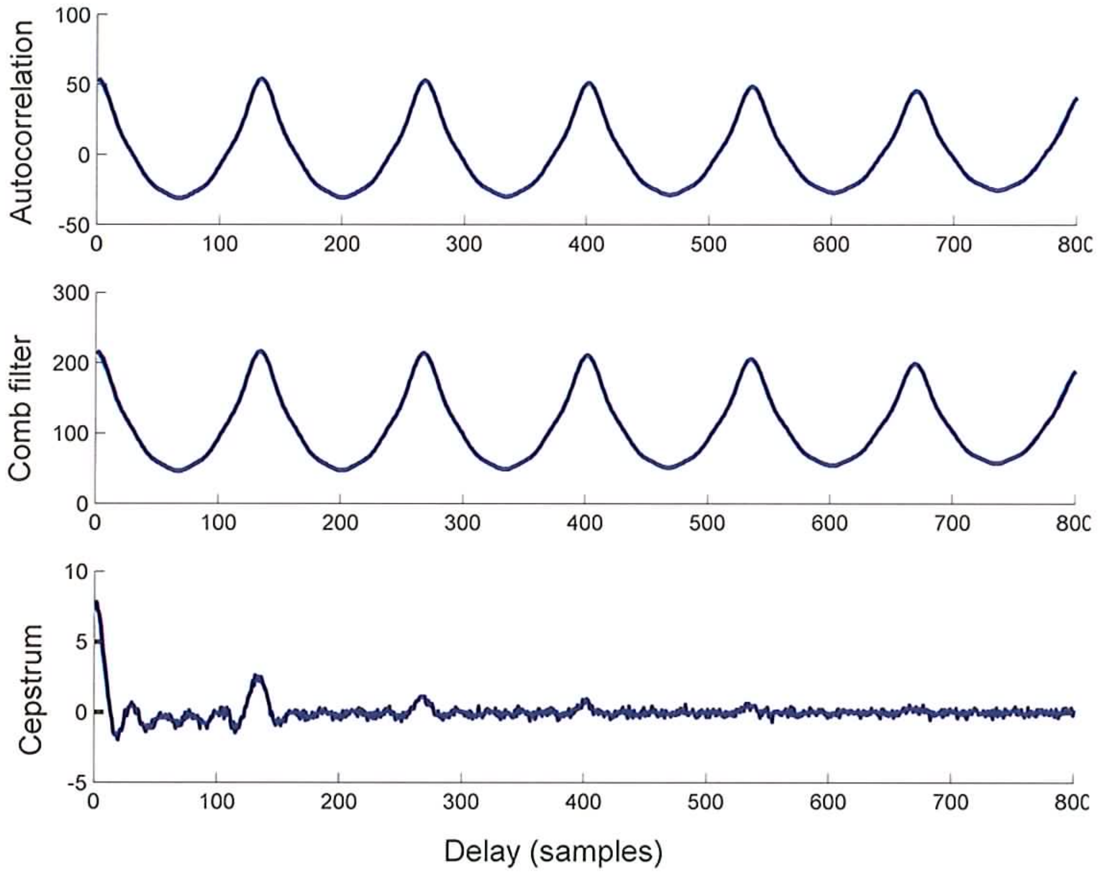

**Figure 5-7: Autocorrelation, comb filter and cepstrum based pitch detection example**

In [Wise '76], the maximum likelihood approach is presented, which is based on analysing the periodicity of the autocorrelation function. The analysis is performed using the equation below for different period $D$ values, and the maximum in the function will correspond to the speech pitch period $D$:

$$MLE(D) = \frac{2D}{N} \sum_{i=1}^{(N-D)/D} r_X(iD) \tag{56}$$

where $N$ is the frame length, $i$ is an integer and $r$ is the autocorrelation function of the signal $x$.

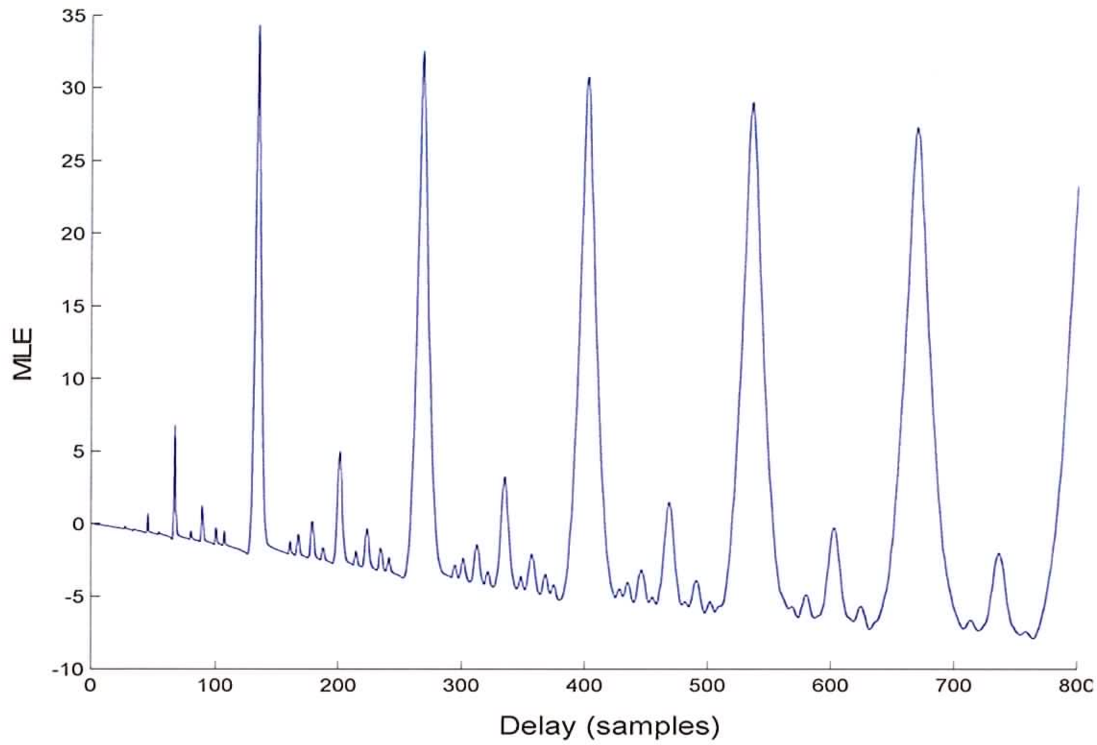

**Figure 5-8: MLE based pitch detection example**

Due to the periodicity of the FIR comb filter impulse response, Equation (56) could also be interpreted as follows:

$$MLE(D) = \frac{2D}{N} \sum_{i=1}^{(N-D)/D} r_X(iD) b_p(iD) \tag{57}$$

where $b_p$ is the impulse response of $(N\text{-}D)/D$ comb filters connected in parallel with the non-zero vector coefficients $b_i = 1$.

As an example, Figure 5-8 depicts a *MLE* based pitch detection of the same synthetic E4 note of Figure 5-1 (top plot). It can be appreciated that the peaks are sharper than in the autocorrelation, cepstrum and comb filter methods. However, other spurious peaks also arise an octave below and in between the multiples of the pitch period.

In [Morgan '97], a spectral weighting function based on [Martin '82]'s method is applied to the *MLE* function of Equation (57) as follows:

$$MLE(P) = \frac{2D}{N} \sum_{i=1}^{(N-D)/D} r_X(iD) b_p(iD) iD^{-0.8} \tag{58}$$

This weighting function attenuates the MLE estimation in delay multiples of the pitch period, which reduces the amount of octave pitch estimation errors. [Martin '82]'s method is reported later in Section 5.2.3, Equation (66).

### 5.2.3 Frequency domain characteristics

Time domain periodicity methods also have a specific frequency domain behaviour, by emphasising the harmonics that are located at frequencies periodically separated along the frequency axis. The standard FIR Comb Filter has its peaks located at multiples of *fs/D*, and by computing the autocorrelation function in the frequency domain, Equation (54) may be expressed as:

$$r_x(D) = \frac{1}{N} \sum_{k=0}^{N-1} |X(k)|^2 \cos(2\pi kD / N) \tag{59}$$

Considering Figure 5-9, where the spectrum of an A4 note signal is depicted, it can be appreciated that by choosing the appropriate delay, the autocorrelation function and the comb filter magnitude response will emphasise the harmonic amplitudes of an A4 note.

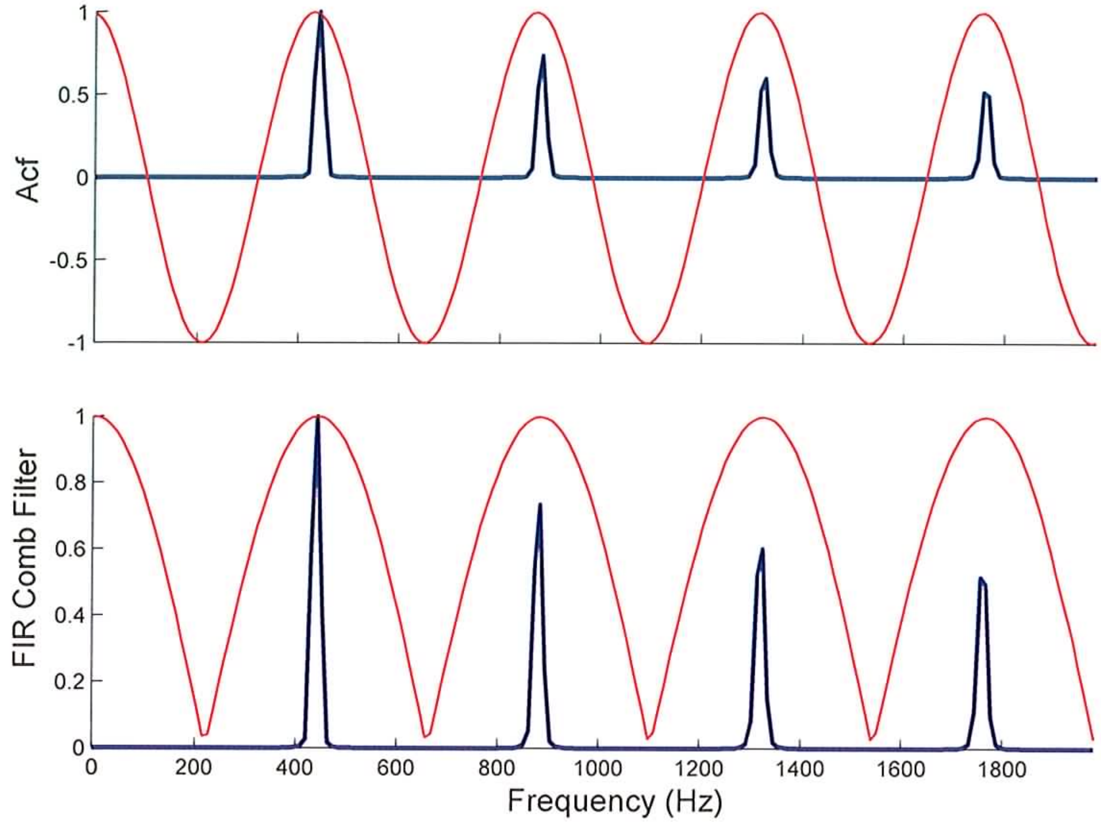

**Figure 5-9: Example of a detection of A4 using the Autocorrelation (*acf*) and FIR com filter pitch 23detection methods**

These methods will perform adequately in a monophonic context. However, their efficiency degrades considerably for the case of polyphonic signal transcription. This is illustrated in Figure 5-10, where a polyphonic signal with notes A4 and C5 playing together and a notch FIR comb filter are depicted. It can be appreciated that even though the Comb Filter extracts the harmonics of A4, the harmonics of C5 are also distorted. Therefore, by using a standard FIR Comb Filter the remaining components of the spectrum after filtering are substantially altered.

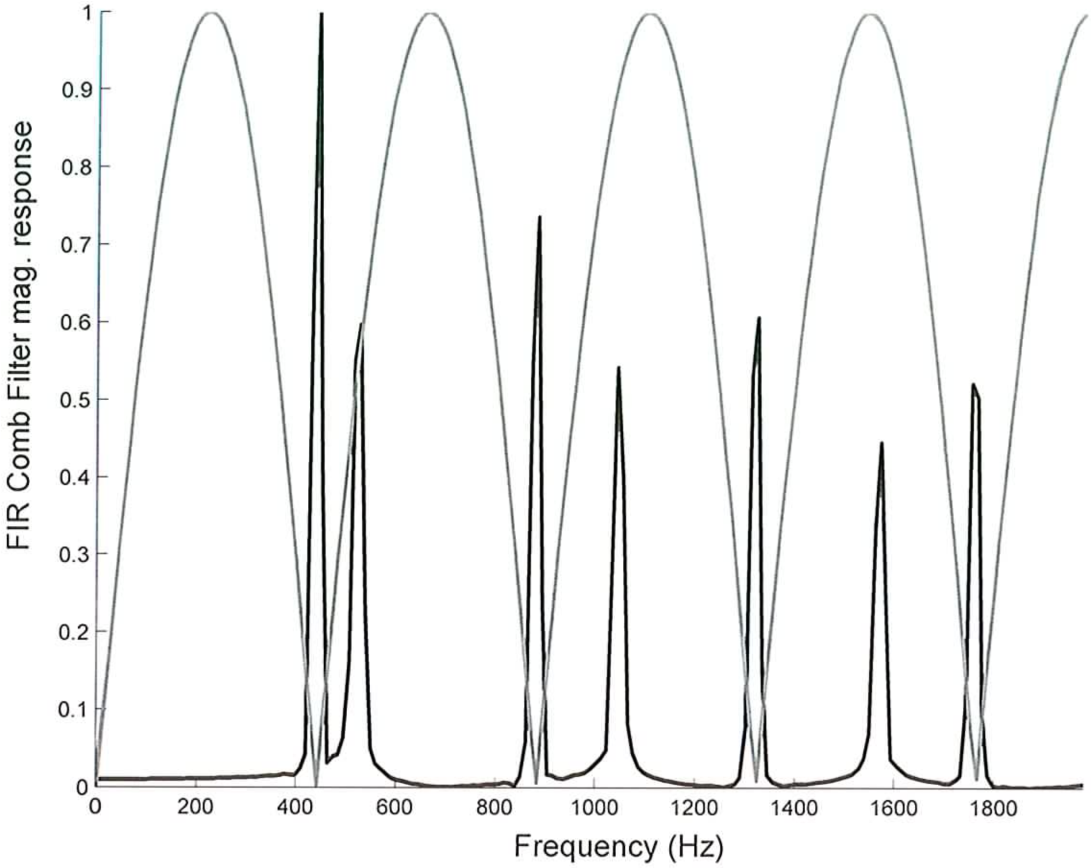

**Figure 5-10: Example of an FIR comb filter detecting A4 when playing with C5**

In a multi-pitch estimation case, it will be of particular interest to extract the harmonics of a given signal whilst leaving the rest of the spectrum unaltered. This is achieved by modifying the filter magnitude response by flattening the filter pass band and increasing the notch region.

A widely utilised approach is to pass the signal through the same filter *H*(*z*) a number of times *p* [Kaiser '77]. Thus, the resulting frequency response *H′*(*z*) is given by:

$$H'(z) = H(z)^{p} \qquad (60)$$

From Figure 5-11, it can be seen that by applying the above equation, low input amplitudes will be attenuated heavily. Thus, the notch band rejection is improved.

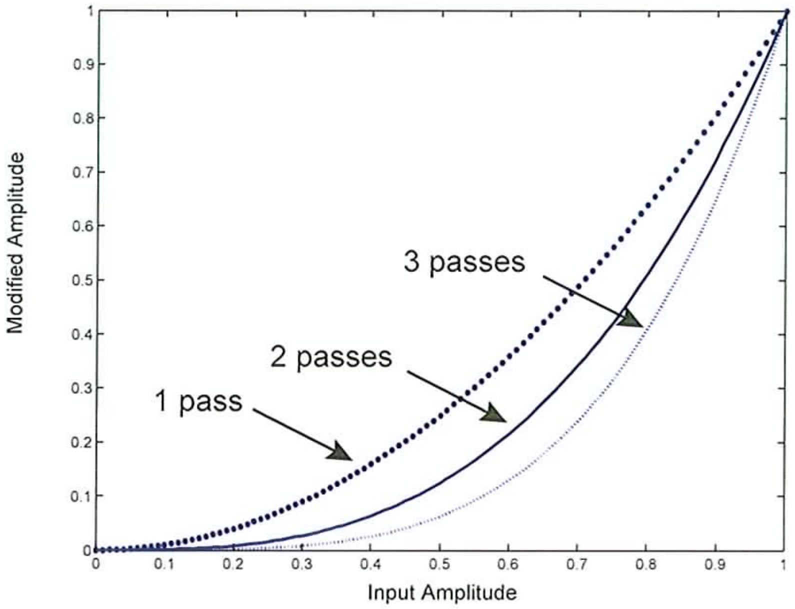

**Figure 5-11: Amplitude change using the same filter**

However, high input amplitudes are also decreased, thus increasing the pass band error. The ideal amplitude change function would be tangential to the points (0,0) and (1,1). In [Kaiser '77], a generalised polynomial is presented, having an $n^{\text{th}}$ and $m^{\text{th}}$ order tangency at points (0,0) and (1,1) respectively. The polynomial is given by:

$$H'(z) = H(z)^{n+1} \sum_{k=0}^{m} \frac{(n+k)!}{n!k!} [1 - H(z)] \tag{61}$$

By way of example, different polynomial orders are depicted in the figure below, where the $H(z)^2$ equation (2 filter passes) is also shown as a reference. For the case of polynomials with $n = m$, the plots have been depicted with a solid line, where it can be appreciated that the tangent flatness improves with the order. If we wish to have a flatter tangent closer to (0,0) than to (1,1) or vice versa, the orders $m$ and $n$ have to be unequal. Two examples illustrating that polynomial type are depicted by the dashed line. The frequency responses of the modified filters utilised in Figure 5-12 are as follows:

$$[H'(z)]_{(m=n=1)} = H^2(3\text{-}2H)$$

$$[H'(z)]_{(m=n=2)} = H^3(10\text{-}15H+6\,H^2)$$

$$[H'(z)]_{(m=3,n=1)} = H^2(10\text{-}20H+15\,H^2\text{-}15\,H^3)$$

$$[H'(z)]_{(m=1,n=3)} = H^4(5\text{-}4H) \qquad (62)$$

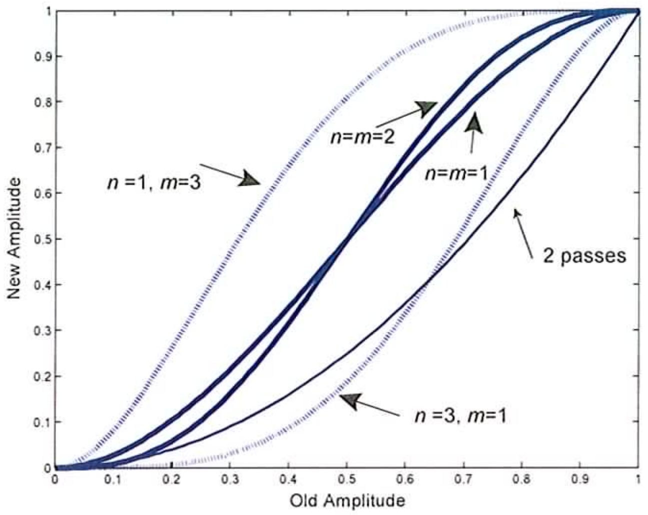

**Figure 5-12: Kaiser's filter sharpening**

Less complex polynomials are proposed by [Valiente '04], which have the following frequency response:

$$H_{1r}(z) = 2H(z)^r - H(z)^{2r} \text{ for } r = 2,3$$

$$H_2(z)= 4H(z)^3 - 3H(z)^4 \qquad (63)$$

An example of the above polynomials are depicted in Figure 5-13, where the polynomials $[H'(z)]_{(m=1,n=3)}$ (labelled $H'(z)$ in the figure) and $H(z)^2$ (labelled "2 passes" in the figure) are also depicted.

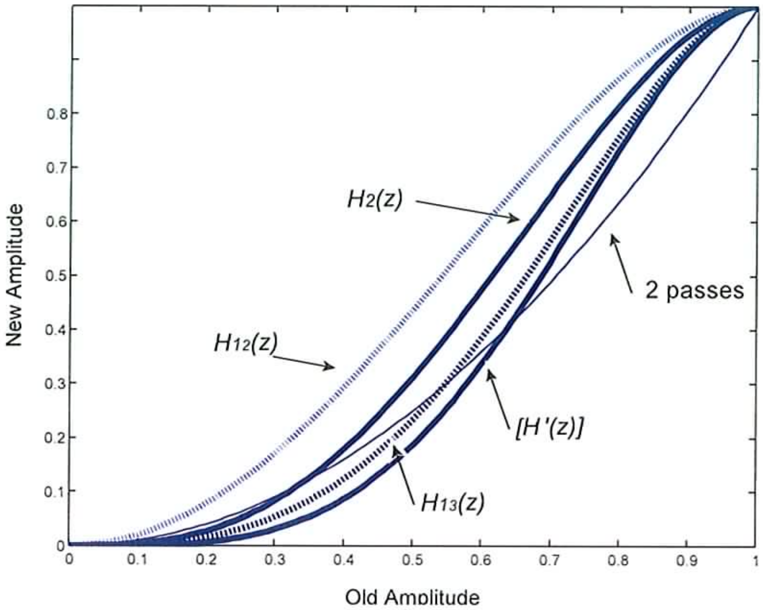

**Figure 5-13: Valiente's filter sharpening**

In [Tadokoro '01], two FIR comb filters are connected in cascade in order to broaden the notch rejection band. The delays of the filters are deviated from an equivalent FIR comb filter *H*(*z*) with delay *D*, a value $\Delta D = 1\%$ of *D*. The frequency response of the modified frequency response is as follows:

$$H'(z) = (1 - z^{-D+\Delta D})(1 - z^{-D-\Delta D}) \tag{64}$$

In Figure 5-14, the FIR comb filter frequency response *H*(*z*) is sharpened by using the above method, and by using Equation (60) with a number of passes $p = 2$. The sharpened frequency responses are depicted by the solid and dotted lines respectively. Tadokoro's method (solid line) has slightly wider notch bands for the first four notches. However, the difference 2**AD* between filter delays, causes spurious peaks in the high notch rejection bands.

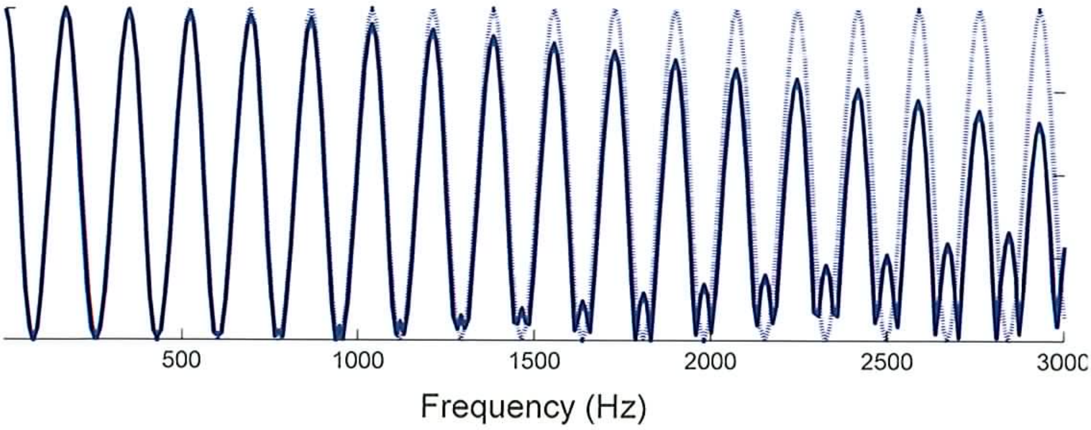

**Figure 5-14: Sharpened frequency response of a comb filter by using Tadokoro's method [Tadokoro '01] (solid line) and Equation (60) method with *p*=2 (dashed line)**

## 5.3 Frequency domain periodicity based methods

Other systems utilise the autocorrelation function in the frequency domain following spectral liftering in order to exploit the periodicity features of the harmonic components of an audio signal [Lahat '87]. The correlation maximum will occur when the delay $D$ matches the fundamental frequency as follows:

$$r_f = \sum_{k=1}^{N-D} X(k) * X(k+D) \tag{65}$$

Kunieda et al. utilise the above method by replacing the magnitude spectrum in the autocorrelation function by its logarithm function, which as in the cepstrum method (Section 5.2) emphasises high frequency components [Kunieda '96].

However, by the use of the autocorrelation function in the frequency domain, harmonic signals will also correlate at multiples of the fundamental frequency.

The use of harmonic frequency patterns to obtain the signal pitch has also been investigated. Martin obtains the fundamental freqency *(f0)* of speech signals by correlating the signal power spectrum and a spectral comb function *C(f)* with decreasing amplitude $i^{-1/c}$,where *c* is an integer [Martin '82]. The spectral comb function *C(f)* for a frequency $f_x$ is as follows:

$$C(f_x, f) = \sum_i i^{-1/c}\delta(if_x - f) \qquad (66)$$

The $i^{-1/c}$ term attenuates the comb peaks with the frequency, which reduces the amount of high octave *f0* detection errors.

In [Brown '92], a spectral representation where the frequency components are logarithmically separated is utilised. Thus, harmonic frequency components are equally spaced in the frequency domain regardless of the fundamental frequency. As an example, the distance between the fundamental and the second harmonic, and between the second and third are log(2) and log(3/2) respectively for all fundamental frequency values. As opposed to Martin's approach (Equation (66)), a common pattern can be built for all *f0* candidates, which when cross-correlated with the spectrum of the signal gives a maximum in the position of the fundamental frequency.

Another frequency pattern approach is presented by [Klapuri '03]. A system that splits the signal into 17 overlapping frequency bands logarithmically distributed between 50 Hz and 6 KHz. Each output band $Z_b$, covers the frequency bins $k$ ε [$k_b$, $k_b$ + $K_b$ -1], where $k_b$ and $K_b$ correspond to the lowest bin of the band and the width of the band respectively.

Then, a weighted vector $L_b(k_f)$ is calculated, where $k_f$ is the fundamental frequency candidate in bins, as follows:

$$L_b(k_f) = \max\left\{c(m,k_f)\sum_{j=0}^{J-1} Z_b(k_b + m + k_f j)\right\} \tag{67}$$

where

$$J(m,k_f) = \frac{(K_B - m)}{k_f}$$
$$c(m,k_f) = \left[\frac{0.75}{J(m,k_f)}\right] + 0.25 \tag{68}$$

and where $c(m, k_f)$ is a weighting function obtained experimentally. Equation (67) is evaluated for different offset $m$ values in order to obtain the $K_f$ that best explains the energy band.

Finally, the bands are summed to obtain the global weight $L(k_f)$.

Frequency domain periodicity methods based on spectral harmonic patterns are sensitive to both high and low octave pitch detection errors. If the harmonics chosen in the pattern are high, the even pattern harmonics will match all the harmonics of the spectrum at half of the fundamental (low octave pitch detection error), which produces the same value in the correlation as that in the true fundamental [Brown '92]. By contrast, at twice the true fundamental, the pattern harmonics will match the even spectrum peaks, which will produce a high pitch detection value [Brown '92].

## 5.4 Auditory models

Whether human auditory perception can separate two different tones depends on the frequency difference between them. For the majority of listeners, if the two tones have a frequency difference less than about 15 Hz, the sound still resembles a single sinusoid whose amplitude varies at regular rates [Howard '01]. These fluctuations are called "beats". It is only when the frequency difference between the two components reaches the "Critical Bandwidth" that the listener is able to perceive smoothly separated tones [Howard '01].

This auditory phenomenon is utilised by [Meddis '91a, b] in his pitch perception model, by first passing the signal through a bank of band pass filters, which models the frequency selectivity of the inner ear. Then the amplitude envelope of each output channel $c$ is obtained, and its periodicity is analysed by utilising the autocorrelation method, $r_C$. Finally, the autocorrelation periodicity estimates are combined across bands to obtain the summary autocorrelation function (SACF), which is given by:

$$s(D) = \sum_{c}^{B} r_c(D) \qquad (69)$$

where $B$ is the number of frequency bands.

In [Meddis '92], the same authors extended their [Meddis '91a, b]'s models to the multipitch case of two simultaneous vowels. The largest peak in the SACF estimation represents the pitch of the prominent vowel. Then, the channels that have a peak in the same delay $D$ are assigned to the same source and are removed. Next, the autocorrelation function of the remaining channels are summed to obtain a residual SACF, which provides the weaker vowel.

De Cheveigne states that the use of a filter bank in order to perform a AMDF analysis in each channel (Equation (52)), provides better results than by obtaining the AMDF of the whole signal [De Cheveigne '91a]. He also shows that the use of the AMDF in high channels provides a higher detection error rate than by applying the AMDF to the whole signal. The former error rate is ameliorated by obtaining the band amplitude envelope prior to the AMDF analysis. However, the same operation in low channels degrades error rates [De Cheveigne '91a].

Another pitch detection method is proposed by [Tolonen '00]. Even though the system does not attempt to simulate the auditory system, it is based on the SACF method and thus included in this section. The model only uses two bands, above and below 1000 Hz, to obtain the SACF. As in [De Cheveigne '91a], the periodicity of the higher band is obtained directly from the band amplitude envelope. In contrast, the lower band periodicity is directly obtained from the signal. Then, the SACF is obtained by summing both functions using Equation (69). In order to avoid detection errors in multiples of the pitch period that typically arise by the use of the autocorrelation function, the SACF function is modified. First, the SACF function is half wave rectified (HWR), and then shifted in time by two (each SACF($D$) will be now located half wave rectified at delay = 2*$D$) and subtracted from the original HWR SACF($D$) signal. The result is once more HWR, which produces the enhanced summary autocorrelation function (ESACF). The same principle applies to other multiples of the pitch period such as 3 times, 4 times and so on.

In [Klapuri '04], an auditory model which combines two different methods for obtaining the periodicity of the resolved and unresolved harmonics is proposed. Resolved harmonics fall separately in different critical bands, and their frequencies can be estimated by analysing the periodicity directly from the signal: *p1*. On the other hand, unresolved harmonics do not fall separately in different bands, and the periodicity of the beats is detected from the amplitude envelope band: *p2*. The overall periodicity of a pitch period candidate is given by:

$$p(D) = p1(D) + p2(D) \tag{70}$$

In order to obtain the signal periodicity, the *acf* of the unitary model is replaced by a harmonic selection method. For each fundamental frequency candidate $f_s/D$, the amplitude of the harmonic number *num* that maximises the following magnitude response region *R* is selected:

$$R(num, D) = num * \left( \frac{fs}{D\text{-}D/2} \cdots \frac{fs}{D + D/2} \right) \tag{71}$$

*p1*(*D*) is obtained by calculating *R*(*num*,*D*) for the first 20 harmonics ($1 \leq num \leq 20$) across the magnitude response bands. Since the degree of harmonic resolvability decreases with the frequency, low harmonics are weighted to a higher degree. The weighted values are then summed together.

In contrast, *p2*(*D*) is calculated directly from the amplitude envelope spectrum, which is obtained by half wave rectifying (HWR) and the LPF in the signal band. HWR generates harmonically related components of around zero frequency [Klapuri '04]. Thus, in the case of two neighbouring harmonics falling in the same band at frequencies *F***n* and *F**(*n*+1), a high magnitude value at the frequency of the beating *F* arises in the band amplitude envelope. The frequency *F* corresponds to the fundamental frequency associated with the neighbouring harmonics. The magnitude value is obtained by again using Equation (71) for *num* = 1 in each band, and then by summing their magnitudes together. In this case, higher bands are weighted more than lower bands. Next, Equation (70) is applied, and the maximum peak in the function *p* is considered to be the first detected pitch.

In order to extend the model to a polyphonic context, the effect of the pitch period *D* is independently cancelled in *p1* and *p2* and the whole process is repeated in an iterative manner until all pitches have been estimated.

## 5.5 Blackboard systems

Blackboard Architectures integrate both signal processing and musical knowledge. The name "Blackboard" arises from the metaphor of a group of experts working together around a physical board to solve a particular problem. The experts follow the evolution of the solution, and contribute to it by modifying the blackboard when their knowledge is required [Martin '96b]. The system is composed of the following modules:

- *The Blackboard,* is a data representation at different hierarchical levels. Data objects are linked forming hypotheses, which are also stored in the board.
- *Knowledge sources,* processing algorithms which manipulate the data (experts). They are independent to each other but can communicate by accessing the same data, or by removing unsupported hypotheses.
- *Scheduler,* which controls the activity of the Knowledge sources.

In [Martin '96b], a system that transcribes piano performances of Bach chorales is presented. The method integrates knowledge sources of garbage collection (elimination of wrong hypothesis), physics (spectral note representation) and music theory (rules governing musical intervals). As can be appreciated in Figure 5-15, the system blackboard follows a five level hierarchy (tracks (amplitude peaks), partials, notes, interval and chords). First, an onset detector is utilised to segment an STFT representation and then, energy peaks of the segments are fed into the blackboard. A refinement of [Martin '96b] is proposed in [Martin '96a]. In order to utilise a front-end that better simulates the auditory system, Ellis's log-lag correlogram is used instead of the STFT [Ellis '96]. This mid-level representation is based on the correlogram, which is the system lowest hierarchical level, and applies a short time autocorrelation *acf* in each band [Slaney '93]. In the log-lag correlogram case, the lags are logarithmically spaced, and the *acf* values for the same lag *D* at different bands are weighted according to the energy of their bands. Then, the weighted *acf*(*D*) values are averaged together to obtain the "periodgram" of the frame. The note hypotheses in the blackboard are now derived from periodicity and onset hypotheses, which are obtained from the peaks of the summary autocorrelation and the signal band envelope respectively.

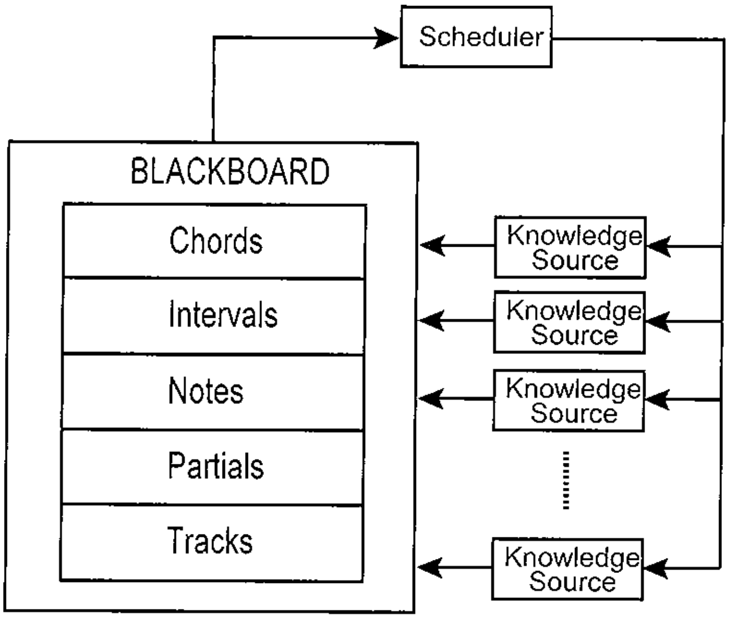

**Figure 5-15: Blackboard system of [Martin '96b]**

Another blackboard system was developed by Kashino et al. [Kashino '95a, b]. The signal is first segmented by using onset times, and each segment is fed into the blackboard which has three levels of hierarchy: frequency components, notes and chords. The knowledge sources hold information of chord progression, probabilities of notes which can occur under a given chord, data of notes played by different instruments, timbre models and Bregman's perceptual rules [Bregman '90]. In order to integrate the knowledge and to process the hypothesis by top-down, bottom-up and temporal processing, a Bayesian probability network is utilised [M. Neal '96].

In [Godsmark '99], a Computational Auditory Scene Analysis (CASA) system that utilises a blackboard architecture in order to integrate Bergman's auditory grouping principles is presented [Bregman '90]. First, a front-end simulating the human auditory system is utilised to form "synchrony strands" based on frequency proximity, temporal continuity and amplitude coherence. The strands, which are the lowest hierarchical level, are combined by the experts to form streams according to Bregman's principles of common onset and offset, temporal and frequency proximity, harmonicity and common frequency modulation. Next, streams are associated to their source of production by their timbre and pitch, which is calculated by simply obtaining the median frequency of the lowest in frequency strand of the stream. Finally, at the highest level of the blackboard, experts are used to identify meter and repeated melodic phrases, which are then fed into lower hierarchical levels to predict new notes.

## 5.6 Data adaptive representations

Data-adaptive representations such as Non–Negative factorization [Lee '01] and Non–Negative sparse coding [Hoyer '02] are not based on prior knowledge of the musical context. Instead, the sources are estimated by learning directly from the data. The methods attempt to approximate an observed data matrix $X_{MxN}$, as a product of two matrixes $A_{MxR}$ and $S_{RxN}$ , where $R \leq M$:

$$X = AS \tag{72}$$

Thus, the $R$ columns of matrix $A$ will correspond to the basis vectors of the decomposition and the $R$ rows of $S$ will contain the source components.

By first analysing the nature of a musical data spectrogram, important observations can be derived. First, the spectrogram contains non-negative data $X$, and the decomposition into non-negative basis functions reflects the intuition that sounds that make up the spectrogram add together. Thus, it can be deduced that if the observed spectrogram $X$ is non-negative, both $A$ and $S$ should be non-negative. In addition, spectrogram data is highly redundant; notes of musical instruments with harmonic structure are inactive most of the time, and are composed of just a few active frequency components. Thus, the hidden components are assumed to be sparse, having a supergaussian probability density function centred at zero with long tails.

In Non–Negative factorization methods, the non-negative matrixes $A$ and $S$ are approximated in order to minimise a cost function, the most basic form of which is given by [Hoyer '02, Lee '01, Smaragdis '03]:

$$C = \|X - AS\|^2 \qquad (73)$$

Equation (73) can be seen as the process of applying the Principal Component Analysis method (PCA) [T. Jolliffe '02] to $X$ with the addition of non-negative constraint. However, by utilising the least squares method in the above equation, it is assumed that the data has a Gaussian distribution, which is not the case in the hidden data of the observed musical spectrogram.

In [Hoyer '02], Hoyer adds the sparseness constraint to Equation (73) making up the Non Negative Sparse Coding method, which is given by:

$$C2 = \|X - AS\|^2 + \lambda \sum S \qquad (74)$$

The above equation attempts to reconstruct the spectrogram as accurately as possible, which is covered by the first term of the equation. In addition, since the sources $S$ are assumed to be sparse and therefore inactive most of the time, Equation (74) attempts to minimise the energy of the second term of the equation by utilising as little energy as possible across the rows of $S$. The parameter $\lambda$ balances the accuracy of the reconstruction with the sparseness of the sources.

In [Lee '01, Smaragdis '03], an alternative cost function is utilised, given by:

$$C3 = \left\| X \otimes \ln\left(\frac{X}{AS}\right) - X + AS \right\|^2 \qquad (75)$$

where $\otimes$ is an element-wise multiplication of matrices.

In [Asari '04], it is demonstrated that Equation (75) is minimised by assuming that the matrix $X$ is generated by a Poisson noise model, which is supergaussian and therefore sparse.

In the case of $X$ being a musical spectrogram, and by utilising the above methods to perform multi-pitch detection, the goal will be to approximate the rows of $S$ as the spectral information of the notes played, and the columns of $A$ as the temporal information of those notes.

This aim is achieved by utilising Equation (73), and choosing the parameter $R$ approximately equal to the number of notes present. However, if $R$ is chosen small, the notes cannot be described. On the other hand, a large $R$ could produce misleading results, since the harmonics of a note could be described as different components. By using Equation (74), the additional components ($R$ – number of notes present) appear as lower energy signals [Smaragdis '03], and by utilising Equation (75) the extra components appear as peaked rows in $S$, with a corresponding spectrum that has very little energy [Smaragdis '03].

## 5.7 Discussion

There are numerous and varied methods that perform pitch detection. A description of the general pitch detection approaches has been described in the Sections 5.2 to 5.6.

Time domain periodicity methods such as comb filter techniques or the autocorrelation function have been widely utilised as monophonic pitch detectors [Brown '93, De Cheveigne '02, Miwa '99a, Noll '64]. An extension to the multi-pitch estimation case is suggested in [Miwa '99b, Miwa '00, Tadokoro '02, Tadokoro '03]. Among these methods, interesting features are provided by [Tadokoro '03], where a parallel configuration of notch FIR filters based on the frequency of the octave 3 notes are used. The filter bank that produces a minimum represents the new detected note. In the same operation, the filter cancels all the harmonics of the detected note, which produces a residual signal to iteratively continue the estimation of the remaining signal pitch. However, due to their frequency domain characteristics, the FIR filters will not be as effective in a multi-pitch estimation context. In this case, a degree of sharpening of the magnitude response is required.

Even though time domain periodicity models have specific frequency domain characteristics, the systems differ from frequency domain periodicity methods. An important difference is found in the type of octave pitch estimation error that the models are prone to. Frequency domain periodicity models are prone to high octave pitch estimation errors. By contrast, time domain periodicity methods are prone to double pitch estimation errors (low octave pitch estimation errors).

Auditory models are also related to periodicity based pitch detection methods. The periodicity of each system band is obtained by applying a time or frequency domain periodicity model to the amplitude envelope of the signal, or directly to the signal. The multi-band configuration reduces the influence of other sources and harmonics when calculating the periodicity of a note within a band. However, in the case of rich polyphonic signals, harmonics belonging to different notes will fall within the same band. Thus, the same problems that standard time and frequency domain periodicity models encounter bandwise will also occur in each band of the auditory model.

The specifications of knowledge based representations are very flexible, and efficiency depends of the knowledge entered and the front-end type utilised. As an example, [Godsmark '99] recreates the human auditory system by using a blackboard system. In [Martin '96b], the note pitches are obtained from a periodicity hypothesis, which relates his system to periodicity based models. An interesting feature of these methods is that the knowledge entered into the system can be utilised to correct pitch estimation errors. A common characteristic of knowledge based representations is the use of an onset detector to segment the signal, which directly relates the methods to the onset detection approaches reviewed in Chapter 4.

Whilst data adaptive representations are promising, they manifest some limitations. Since there is no robust way of choosing $R$, it is difficult to incorporate them into a fully automated transcription system. In addition, the system does not identify individual notes, but rather events. Therefore, if a chord is played several times, and the single notes that comprise the chord are not played outside the chord, there is not further evidence that the chord is not just a single entity.

## 5.8 Conclusions

In this review chapter, the main existing pitch detection approaches have been reviewed, which include methods based on detecting the time domain or frequency domain periodicity of harmonic signals (Sections 5.2 and 5.3), modelling the auditory system (Section 5.4), incorporating knowledge of the signal (Section 5.5), or data representations (Section 5.6). The different methods have been discussed in Section 5.7, which outlines the degree of overlap between the different pitch detection approaches, and the difficulties in classifying them.

As introduced in Chapter 2, current Irish traditional music can be played solo, in unison or even utilising harmony. When it is played solo, traditional Irish players frequently utilise ornamentation to embellish the melody. Ornamentation is a very important musical feature in Irish traditional music, and it is understood differently to classical music. A system that detects the different types of ornamentation within Irish traditional music has not yet been implemented. Since ornamentation in Irish traditional music possesses inherent musical rules, the exercise of integrating knowledge into the system as in blackboard approaches could be an advantage. Ornamentation events occur for a short duration of time before the note they ornament. Thus, in order to transcribe the ornament, the note and ornamentation event time should be obtained prior to the ornamentation pitch analysis by using an onset detector. Consequently, a robust onset detection system will be first required, which directly associates pitch and onset detection problems.

In Chapter 4, an extended literature review of onset detection methods has been provided. The main problems encountered by the existing onset detection approaches are related to signal modulations, legato passages and also in the detection of slow onsets. The majority of Irish traditional music instruments such as the tin whistle, fiddle, flute, concertina and uilleann pipe have a slow onset. Thus, a system that utilises onset information to perform pitch detection within Irish traditional music has to first address all these issues.

In the following chapter, an onset detection system which focuses on the characteristics of the tin whistle within Irish traditional music is introduced [Gainza '04c], whose specifications are derived from the discussion and conclusions documented in Chapter 4. In Chapter 7, the onset detection results provided by this onset detection system are combined with ornamentation theory in order to detect the ornamentation played by the tin whistle, including cuts, strikes, rolls and cranns. The development of this novel ornamentation system forms the basis of Contribution 2.

As introduced in Chapter 2, Irish traditional music has been historically played in unison. In this case, the onset detector presented in Chapter 8, which deals with slow onset instruments modulated in amplitude and frequency, working conjointly with one of the periodicity based pitch detectors reviewed in this chapter, may be used to transcribe the notes.

In recent years, harmonic accompaniment has been added to Irish traditional music. As discussed in the previous section, periodicity based pitch detection methods are less efficient when there is more than one source present in the signal or in the frequency band. In this case, a degree of sharpening of the magnitude response is required. In Chapter 9, a multi-pitch estimator previously introduced in Section 1.2.1 as Contribution 4 is presented. The system is based on the [Tadokoro '03] method, which provides flexibility to build a comb filter per note of the considered pitch detection range. This will also permit connecting a key/mode detector to the pitch detection system, which will reduce the number of filters/notes to be considered of the detection. As in [Tadokoro '03], the model also utilises comb filters, whose harmonic type of magnitude response provides a very useful feature for dealing with harmonic signals. This observation is strengthened by the results provided by the onset detector presented in Chapter 8 [Gainza '05b], which is a robust method on detecting onsets within Irish traditional music. In the multi-pitch estimation case, the FIR comb filters of [Tadokoro '03] are replaced another type of comb filter, which alters the remaining spectrum after filtering to a lesser degree.

In the following chapter, the first onset detector proposed is implemented, customising the system according to the characteristics of the tin whistle.

# 6 Onset Detection System applied to the Tin Whistle (ODTW)

## 6.1 Introduction

In Chapter 4, the main methods that perform onset detection were reviewed. The difficulties that existing onset detection approaches encounter when detecting slow onsets, dealing with signals modulated in amplitude and frequency, or during fast passages such as legato, have also been discussed. One example of a slow onset instrument is the Irish tin whistle. This instrument frequently produces amplitude and frequency modulations, and its legato nature of playing sets a challenging context for existing onset detectors. As documented in Chapter 2, the tin whistle is an important instrument in Irish traditional music, and the development of a system capable of detecting its onsets is investigated in this chapter. The method presented focuses on different aspects of the tin whistle within Irish traditional music [Gainza '04c] to customise the onset detection system, representing Contribution 1 in Section 1.2.1.

Onset detection and pitch detection are crucial tasks in music transcription. In Chapter 7, an extension of the presented onset detector for detecting single and multi-note ornamentation transcription is presented [Gainza '04a].

This chapter is subdivided into the following parts: Section 6.2 describes the different blocks that comprise the onset detection system, and details different system configurations within which each block can operate. In order to optimise the accuracy of the presented onset detector, the configurations introduced in Section 6.2 are evaluated in Section 6.3 for a wide range of system parameters. In addition, the performance of the method is compared against existing approaches, which are also evaluated for different system parameters. A discussion of the results obtained in Section 6.3 is presented in Section 6.4, which leads to the conclusions documented in Section 6.5.

## 6.2 Onset detection system

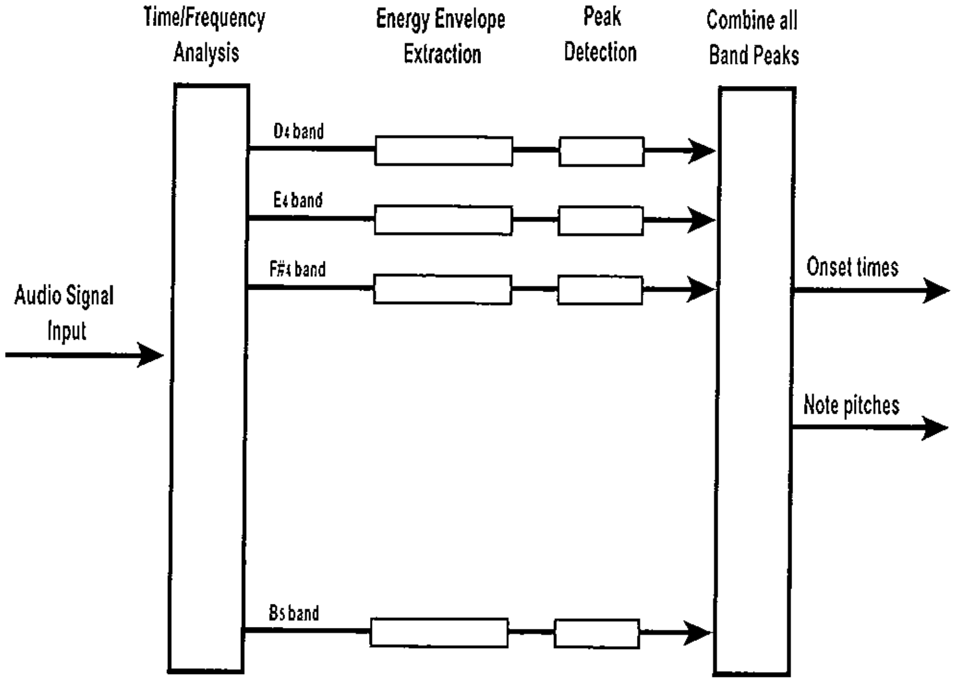

**Figure 6-1: ODTW system overview**

This section describes the different blocks of the proposed onset detector (Figure 6-1). A time - frequency analysis is first required, which splits the signal into different frequency bands. The energy envelope is calculated and smoothed for every band. Peaks greater than a band dependent threshold in the first derivative function of the smoothed energy envelope will be considered as onset candidates. Finally, all band peaks are combined to obtain the correct onset times.

In order to optimise the onset detection accuracy, different configurations of each block of the system are investigated. The performance of these configurations is evaluated in Section 6.3 for different system parameters.

### 6.2.1 Time-Frequency Analysis

The audio signal is first sampled at *fs* = 44100 Hz. Next, the frequency evolution over time is obtained using the Short Time Fourier Transform (STFT), which is calculated using a Hanning window and an FFT length $N = 4096$. The STFT is given by:

$$X(n,k) = \sum_{m=0}^{L-1} x(m+nH)w(m)e^{-j(2\pi \; N)km} \tag{76}$$

where *w(m)* is the window that selects an $L$ length block from the input signal $x(m)$, $n$ is the frame number and $H$ is the hop length in samples.

Figure 6-2 depicts a rapid legato transition between close notes (G4 and F4# in this example). It can be seen that the F4# onset note does not produce an energy increase, and will be missed by existing energy based onset detectors if both successive notes G4 and F4# fall within the same band.

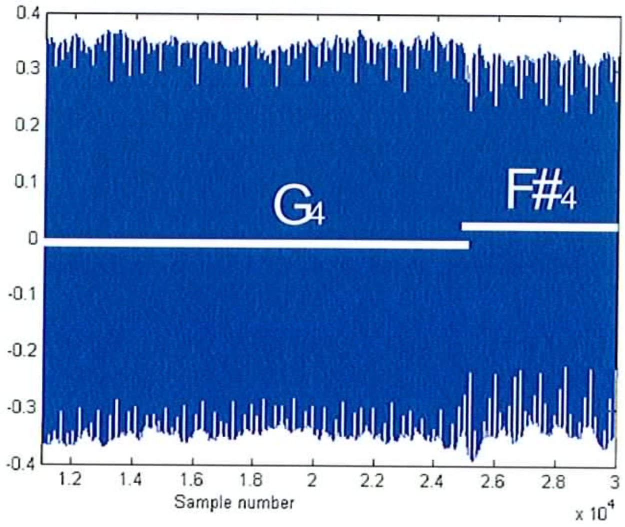

**Figure 6-2: $G_4$-$F\#_4$ tin whistle notes transition example**

Section 2.3 mentions that playing the tin whistle by half covering the holes is not practical in many musical situations. However, the tin whistle can play a wide range of modes commonly utilised in Irish traditional music without using half covering (described in Section 2.3). The notes that can be played with full covering of the holes (which allows the instrument to play in those modes) are shown in Table 6-1.

| ***Octave 4*** | | | | | | | | ***Octave 5*** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***D*** | ***E*** | ***F#*** | ***G*** | ***A*** | ***B*** | ***C*** | ***C#*** | ***D*** | ***E*** | ***F#*** | ***G*** | ***A*** | ***B*** |

**Table 6-1: Full covering notes for the D tin whistle**

In order to overcome the problem of having more than one note falling within the same band, each frame is filtered using a bank of 14 band pass filters (each one representing a different note from Table 6-1). Thus, the problem inherent in legato playing is reduced, since each note will arise in its corresponding band. As an example, the energy of G4 and F#4 in Figure 6-2 will fall in separate bands. From Table 6-1, it can be seen that C and C# are not utilised in the higher register. These notes are not commonly used by tin whistle players [Larsen '03], which is corroborated by [Duggan '06a] who states that these notes sound shrill and are thus avoided by tin whistle players. Each band is centred at the frequency of its corresponding note (Table 6-1) according to the equal tempered scale [Lindley '06]. The width of each band $i$ starts at the middle distance between the centre frequency of the precedent band $i$ -1 and the current band $i$, and finishes at the middle distance between the centre frequency of current band $i$ and the next band $i + 1$. In order to generate D4 and B5 bandwidths in Table 6-1, the frequencies of C#4 and C6 are utilised as the frequency of the preceding and next band respectively.

A windowed signal frequency component is highly dependent upon the choice of the frame length $L$. For example, a Hanning window has a main lobe width equal to 4 bins, where in Equation (76) $N = L$ [Proakis '95]. Thus, for $L = 4096$, the bin width and the main lobe width will be equal to $fs/N$ = 44100/4096 = 10.77 Hz, and 43.01 Hz respectively. As an example, if F4# is played perfectly tuned to the equal tempered scale, the main lobe of F4#, $ML_{\mathrm{F4\#}}$, will be located at the frequency region $[ML_{\mathrm{F4\#},}]_{L=4096}$ = [740 – (43.01/2): 740 + (43.01/2)] = [718.49: 761.5]. In the same manner, the main lobe width of G4 will cover the frequency region $[ML_{\mathrm{G4}}]_{L=4096}$ = [762.5: 805.5]. However, the time resolution provided by using such frame length is not adequate for onset time estimation, which requires a higher degree of accuracy.

By using smaller frames, e.g., $L$ = 1024 samples, the main lobe width is now equal to 4*($N$/$L$) = 16 bins, which corresponds to approximately 172 Hz. In this case, the F4# and G4 main lobe widths will be located at the frequency regions $[ML_{\mathrm{F4\#},}]_{L=1024}$ = [567.68,912.32] and $[ML_{\mathrm{G4}}]_{L=1024}$ = [611.68, 956.32] respectively. In this case, the energy of a note will be present in more than one band. This shows that the choice of $L$ is crucial in signal analysis, and determines the compromise between the frequency resolution required to separate very close notes in different bands, and the time resolution required to provide accurate onset time estimation. In Section 6.3, the impact that the choice of $L$ and $H$ has in the onset detector is investigated.

### 6.2.2 Energy Envelope and signal smoothing

The energy is calculated in each band for each frame using:

$$E_{(i,n)} = \sum_{k_i=1}^{l_i} \left\{ |X_i(k_i, n)|^2 \right\} \tag{77}$$

where $X_i$ is the filter output of band $i$, $k_i$ is $i$'s $k^{\text{th}}$ frequency bin number and $l_i$, is the band i's length in frequency bins.

As an example, an A4-G4-F#4 note transition is depicted in Figure 6-3, plot A. The energy envelopes of the bands F#4, G4 and A4 are depicted in plots B, C and D respectively, where it can be seen that each note falls in a separate band.

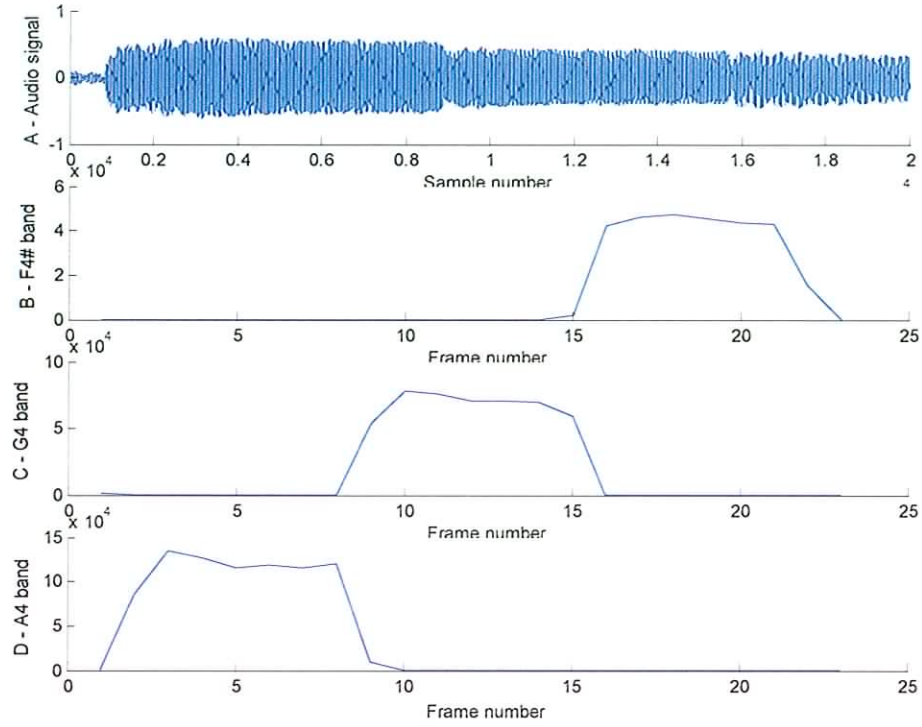

**Figure 6-3: A4-G4-F#4 tin whistle note transition example (plot A). Plots B, C and D depict the energy envelopes of bands F#4, G4 and A4 respectively**

[Scheirer '98] and [Klapuri '99] smooth the amplitude envelope by convolving the average energy signal with a Half Hanning window (Figure 4-5). However, the lack of symmetry of the Half Hanning window produces a non linear phase window, which considerably alters the envelope of the time domain signal. This is illustrated in Figure 6-4, where smoothed versions of the onset detection function depicted in plot A, are depicted in plots B and C by using the first and the second half of a 200 ms Hanning window respectively. It can be seen that by using the second half (plot C), the peak located at sample 50 in plot A is distorted at the peak decay, and by using the first half (plot B) the resulting smoothed signal will be distorted at the attack peak.

A possible solution would be the use of a full Hanning window, which has a symmetric shape and a linear phase. However, the filter kernel would be too long, delaying the smoothed signal considerably. A more accurate solution will be to perform zero-phase filtering by using a low order IIR filter, which processes the input signal in both the forward and reverse directions. This is illustrated in the plot D of Figure 6-4, where a 3rd order IIR filter with cutoff frequency *fc* = 0.3*(*fs*/2) has been utilised to smooth the plot A signal. It can be seen that the IIR filter deals with peaks and decays in a similar manner, and that the resulting smoothed signal more accurately preserves the signal envelope shape.

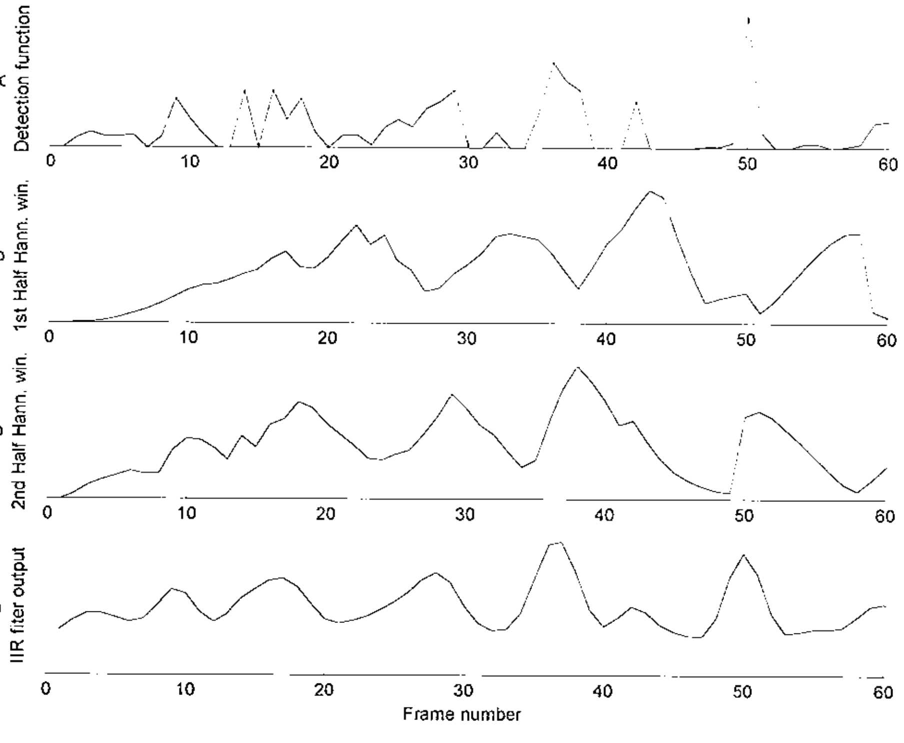

**Figure 6-4: Smoothed versions of the noisy onset detection function depicted in plot A. The techniques used are labelled in the *y* axis of plots B, C and D respectively.**

By using longer Hanning windows or lower cutoffs, the smoothness of the signal will increase. The effect that this parameter has in the onset detection accuracy is evaluated in Section 6.3.

### 6.2.3 Peak Picking and Thresholding

As in [Scheirer '98], the first order difference of the energy envelope is calculated for each band, and peaks that reach a predetermined threshold will be considered as onset candidates. As discussed in Section 4.2, the first order difference function delays the onset estimation time. However, the method is also less prone to spurious detections than the relative difference function of [Klapuri '98], which is an important characteristic in signals modulated in amplitude, such as tin whistle signals.

An onset candidate will be detected if:

$$E_{(i,n)} - E_{(i,n-1)} > T_i \qquad (78)$$

where $E_{(i,n)}$ denotes the smoothed energy envelope.

Other multi-band energy based approaches such as [Klapuri '99, Scheirer '98] use the same threshold for every band. However, this is not adequate for wind instruments such as the tin whistle, where strong amplitude modulations in high bands can have similar peak values as onset peaks in low bands. In addition, different notes have a tendency to be played at a different blowing pressure. In this section, four different thresholds are considered in order to obtain a different threshold per band in the ODTW. One of the methods was developed by [Duxbury '02], and is denoted as ***Thres1***. The remaining three methods presented in this section are novel, which are denoted as ***Thres2***, ***Thres3*** and

***Thres4.*** The development of these thresholding methods has been introduced in Section 1.2.1 as part of Contribution 1, and their performance is investigated in Section 6.3. The methods are classified according to the signal properties they exploit: statistical, acoustical, or a combination of both acoustical and statistical properties.

- **Statistical based thresholds**

Two different statistical based thresholds, *Thres1* and *Thres2,* are considered to pick the onset candidates from the band onset detection function. The methods are introduced as follows:

- ***Thres1***: The threshold is set at the **second derivative of the histogram** of the band detection function [Duxbury '02]. This method was presented in Section 4.4.
- ***Thres2***: A novel method based on the **standard deviation** (*std*) [Pal '05] is presented. The *std* provides an estimation of how the values of a signal $x$ deviate from the mean of the signal, and is given by:

$$std(x) = \sqrt{\frac{1}{S}\sum_{j=1}^{S}(x_j - \bar{x})^2} \qquad (79)$$

where $\bar{x}$ is the mean of $x$, and $S$ is the number of samples.

When an onset occurs in a band, the value in the detection function is significantly prominent. Consequently, the onset peak value deviates from the band mean more than the band standard deviation. Thus, the band $i$ threshold will be given by:

$$T_i > xi + std(xi) \qquad (80)$$

where $xi$ is the onset detection function of a band $i$.

- **Acoustic properties based threshold :*Thres3***

Each note of a wind instrument has a different pressure range within which the note will sound satisfactory; this range increases with the frequency [Martin '94]. Martin states that the usual practice for recorder players is to use a blowing pressure proportional to the note frequency, thus the pressure increases by a factor of 2 for an octave jump. We can then conclude that as with the note frequency, the general blowing pressure for different notes is spaced logarithmically. The same principle is applied to the tin whistle, due to its acoustic similarity with the recorder [Gainza '04c].

In both cases, the threshold should also be proportional to the frequency and will have a logarithmic spacing. Then, according to [Martin '94]'s theory, a novel band threshold is implemented, which for a band *i* will be given by:

$$T_i = T * 2^{\frac{s}{12}} \tag{81}$$

where $T$ is the reference threshold required for the band of a given note $x$, and $s$ is the semitone separation between the note in the $i$ band and the reference note $x$.

In order to obtain the band energy $E_i$, $X_i$ in Equation (77) is squared. Thus, the threshold will also have to be squared as follows:

$$T_i = T_i^2 \tag{82}$$

- **Combining statistical and acoustical properties: *Thres4***

Even though the acoustic based approach provides a different threshold per band, this method also requires user input to set the reference band threshold $T$. In order to overcome this limitation, the novel acoustic and statistical based methods are combined into one unique method. The reference threshold $T$ of Equation (81) will be set by using the statistical method *Thres2*. Then, the remaining band thresholds are set according to the acoustic based method *Thres3*. In order to ensure setting the reference threshold *T* in a band that contains an active note, the reference threshold *T* is set in the band that contains more energy.

In order to set the band thresholds, *Thres3* and *Thres4* utilise acoustic characteristics of the tin whistle such as the expected blowing pressure per note is utilised. This justifies the use of an energy band approach to produce band onset detections, which provides an indication of the blowing pressure per note.

### 6.2.4 Combing the peak bands

Onset candidates in each band are combined and sorted in time (frame number). The methods described in Section 6.2.3 set the band thresholds according to the energy band envelope. Thus, in the case of not having any note active in a band, the energy band envelope may have been obtained from the energy spread of active notes in adjacent bands or from signal noise. In this case, Section 6.2.3's methods will produce a very small threshold value. In order to avoid spurious onset detections due to low threshold values in low energy content bands, only bands containing active notes should be considered.

In the ODTW, the maximum envelope value of bands that do not contain active notes will be considerably lower than in bands that contain active notes. Thus, if the increase between successive maximum band envelope values is unusually large: $me(i) - me(i\text{-}1) > 10^* me(i\text{-}1)$, where $me(i)$ is the maximum envelope value of a band $i$, it will be assumed that there is not any note active in band $i$-1.

In the same manner: if $me(i\text{-}1)$ - $me(i)$ > 10* $me(i\text{-}1)$, it will be assumed that there is no note active in band $i$. In both cases, the factor 10 value is an experimental value obtained through testing.

Finally, as in [Bello '03, Duxbury '02, Duxbury '03a, Klapuri '99], a sliding window *win* centred at each onset candidate is applied. The most prominent candidate is maintained, while the remaining onset candidates are assumed to belong to the same onset and so are discarded. This parameter (*win*) is also utilised in Section 6.3 to determine the tolerance in the onset detection accuracy between the target onset and the onset candidate. The impact of the choice of *win* in the test results is evaluated in Section 6.3.

## 6.3 Results

In order to evaluate the performance of the presented approach, a hand labelled test signal database of 493 tin whistle onsets belonging to 11 excerpts of Irish traditional music tunes sampled at $fs$ = 44100 Hz was first created. Three different players produced the tunes that comprise the test. The tunes have been selected from commercial CD recordings, as well as recorded during informal live sessions. A wide range of traditional music tune types is represented in the database, covering reels, slip jigs, single jigs, double jigs and slow airs. The players produced amplitude and frequency modulations and utilise a large variety of ornamentation types: cuts, strikes, rolls, crams, vibratos and slides.

First, the performance of the ODTW is evaluated in Section 6.3.1 by using the different configurations and parameters described in Section 6.2. Then, a comparative analysis is performed in Section 6.3.2 by comparing the ODTW against existing onset detection methods. These approaches do not differentiate between ornamentation and note events.

Thus, in order to perform a fair comparison against the existing methods, ornamentation events are considered as onsets, even though traditional Irish music considers single-note ornamentation events as part of the onset [Larsen '03].

The accuracy of each onset detection method, which is denoted as *acc*, is calculated by using the following equation [Klapuri '99]:

$$acc = \frac{total - FN - FP}{total} 100\% \qquad (83)$$

where *total*, *FN*, and *FP* are the total number of onsets, false negatives (undetected), and false positives (spurious) respectively.

An FFT length equal to 4096 samples is utilised for all the calculations. A discussion of the results obtained in this section is provided in Section 6.4.

### 6.3.1 Evaluation of ODTW for a wide range of System Parameters

It has been shown in Section 6.2 that different system parameters and configurations can be utilised in each block of the ODTW; these are summarised as follows:

- *TFP* (*Time Frequency Parameters*): as introduced in Section 6.2.1, the time-frequency analysis is highly dependent on the pair of parameters *L* and *H*.
- *fc*: in Section 6.2.2, the energy envelope is smoothed by using an LPF, whose degree of smoothness depends on the choice of *fc*.
- *TM* (*Thresholding method*): in order to pick the onset candidates from the onset detection function, different methods for setting up the band threshold have been introduced in Section 6.2.3.

- *win:* in order to combine the band peaks, a window with length equal to *win* is applied in Section 6.2.4. In addition, onset candidates falling within a window length *win* centred at the target onset location are considered correct detections. Thus, *win* also determines the degree of tolerance in the detection.

In order to optimise the accuracy of the ODTW, the system is evaluated by varying the above system parameters and configurations. Two different tests are performed: the first test, which is denoted as TEST 1, evaluates the entire test material by using different *TFT* and *TM*, where *fc* and *win* are assigned a constant value (as is usually the case in existing methods [Bello '03, Duxbury '02, Duxbury '03b, Klapuri '99]). The second test, TEST 2, first configures the system with the configurations and system parameters that provide the best results in TEST 1. Then, the entire test material is evaluated for different *fc* and *win* values.

### TEST 1: Evaluation of *TFT* and *TM*.

The ODTW is evaluated for the entire signal database by utilising the configurations and system parameters described by *TFT* and *TM*, which are given by:

- *L/H*: the pair of system parameters *L/H* equal to 512/1024, 1024/2048, 512/2048 are considered.
- *TM*: only automatic thresholds are considered to evaluate the ODTW. From Section 6.2.3, the three methods that meet this requirement are the novel methods ***Thres2*** and ***Thres4*** and the existing method ***Thres1*** [Duxbury '02]. As

in [Beardah '95], the probability distribution function (*pdf*) in *Thres1* is generated by smoothing the histogram with a triangular kernel.

The system parameters *fc* and *win* are kept constant, where values are obtained as follows:

- *fc:* the detection functions of all methods are smoothed by using a 3rd order IIR filter with *fc* equal to 0.4*(*fs/2*).
- *win*: onset candidates falling within a window *win* equal to 50 ms window centred at the target onset location are considered correct detections.

The accuracy results *acc* (Equation (83)) are shown in Table 6-2 for the above mentioned parameters and configurations. As in [Bello '03], the percentage of good positives (good detections), *pGP*, is calculated by dividing the number of good positives by the total of onsets in the hand labelled database. The percentage of *FN* and *FP*, *pFN* and *pFP*, are calculated by dividing *FP* and *FN* by the number of onsets picked from the detection function [Bello '03].

| | *Thres1* | | | | *Thres2* | | | | *Thres4* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *H / L* | *pGP* | *pFN* | *pFP* | *acc* | *pGP* | *pFN* | *pFP* | *acc* | *pGP* | *pFN* | *pFP* | *acc* |
| **512/1024** | 84.99 | 10.10 | 42.84 | 21.30 | 81.34 | 18.59 | 18.99 | 62.27 | 74.24 | 29.47 | 15.08 | 61.05 |
| **512/2048** | 82.76 | 11.74 | 43.65 | 18.66 | 80.32 | 19.92 | 18.69 | 61.87 | 74.44 | 28.77 | 16.21 | 60.04 |
| **1024/2048** | 87.83 | 9.19 | 33.69 | 43.20 | 81.74 | 19.23 | 13.89 | **68.56** | 73.02 | 32.68 | 11.55 | 63.49 |

**Table 6-2:** ***pGP, pFN, pFP*** **and** ***acc*** **results (in %) obtained by evaluating the ODTW for different** ***H/L*** **pairs and the thresholding methods** ***Thres1*****,** ***Thres2*** **and** ***Thres4*****.**

### TEST 2: Evaluation of *fc* and *win*

In TEST 1, onset candidates falling within a 50 ms window centred at the target onset location are considered correct detections. In addition, an *fc* equal to 0.4*(*fs/2*) is utilised to smooth the onset detection signal. However the tin whistle is commonly played by using ornamentations such as cuts and strikes, which can last less than 50 ms before the onset note time. In this case, both ornamentation and note onsets will fall within the same window and only the strongest peak will be maintained. LPFs smooth the onset detection signal, thus avoiding multiple detections of the same onset. However, a very low *fc* can merge the successive ornamentation and note onsets into one unique peak.

In order to obtain the best pair of *win* and *fc*, the ODTW is first configured with the best performing set of parameters, *TFT* and *TM*, obtained by evaluating TEST 1. From Table 6-2, the best *acc* result is equal to 68.56 (shown in bold). The parameters that provide this *acc* result are: $H = 1024$, $L = 2048$ and the thresholding method *Thres2*.

Then, the ODTW is evaluated for the entire test material by utilising the system parameters *fc* and *win*, which are described as follows:

- Different *win* values scaled by steps of 5ms within the [20:50ms] range.
- Different *fc* values scaled by steps of 0.02*(*fs/2*) within the [0.2:0.7]*(*fs/2*) Hz range.

The results are depicted in Figure 6-5, where each of the lines represents the *acc* results obtained by evaluating the ODTW for a fixed *win* value within the entire *fc* range. The best *acc* result is equal to 71.60, and it is obtained by using *fc* equal to 0.28.

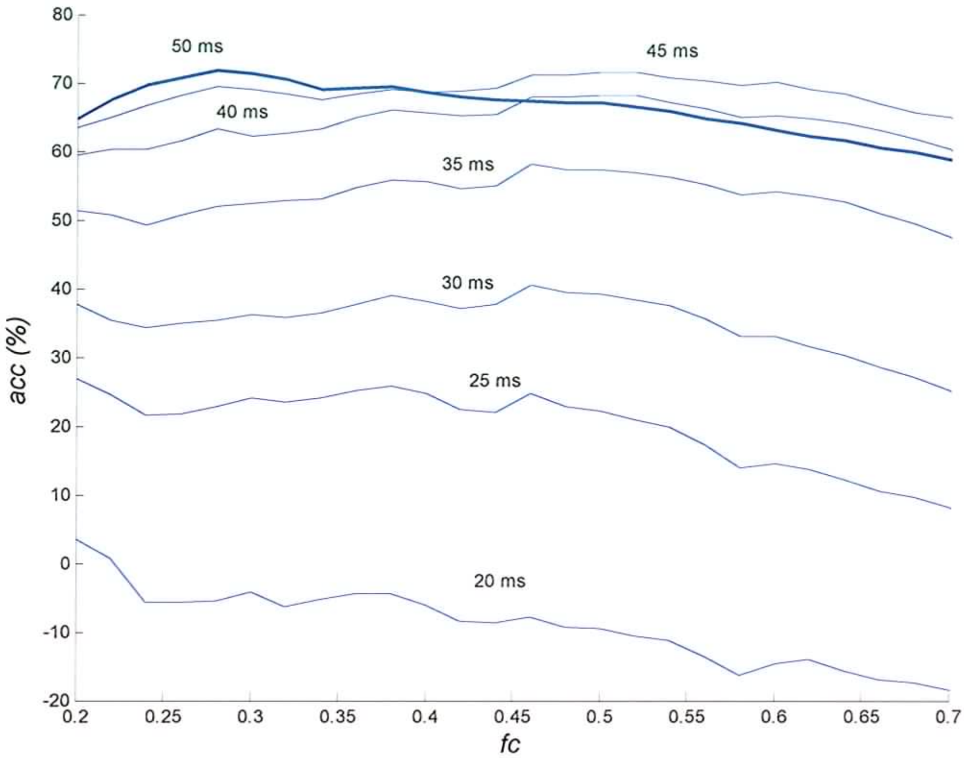

**Figure 6-5:** ***acc*** **results obtained (*y* axis) by using different *fc* values (*x* axis range) and *win* values. Each line of results is obtained by using the *win* value that labels its corresponding line.**

### 6.3.2 Comparison between Methods

A comparative analysis of the ODTW against existing onset detection methods is performed in this section. First, the existing methods are evaluated for the entire signal database. Then, the results are compared against the results obtained by ODTW in Section 6.3.1. The diverse existing systems utilised are: the energy based approach spectral difference method [Duxbury '02], the phase based method [Bello '03], and the combined energy and phase approach complex based method [Duxbury '03b]. In order to investigate the impact of high frequencies on detecting tin whistle onsets, a method based on [Masri '96b] is utilised. The method obtains the first order difference of the High Frequency Content (Equation (16)), which is denoted as *d*(HFC) As in Section 6.3.1 for the ODTW, existing onset detection systems are also evaluated by using the same testing methodology. First, TEST 1 evaluates the entire test material using different *TFT* and *TM*. Then, TEST 2 evaluates the entire test for different *fc* and *win*.

### TEST 1: Evaluation of *TFT* and *TM*

When evaluating the existing onset detection methods, the median filter approach of [Bello '03] is now utilised instead of *Thres4*, which is denoted as *Thres5*. As indicated in Section 4.3, this method compares the onset detection results obtained by iteratively applying a median filter to the onset detection function against a hand labelled database. For each method, the required $\delta$ parameter (Equation (39)) is scaled by steps of 0.01 from 0 to 1, thus running the algorithm 100 times. The $\delta$ that produces the best results in the normalised detection function is obtained from the closest point to the top left corner of a figure that displays *pGP* against *pFP* for all $\delta$ values [Bello '03]. As an example, Figure 6-6 depicts the *pGP* vs. *pFP* plot for all the existing onset detection methods considered in this section for a *H* / *L* pair equal to 512/1024 samples. The best $\delta$ value per method using this *H* / *L* pair is obtained at $\delta_{complex} = 0.04$, $\delta_{HFC} = 0.01$, $\delta_{specdiff} = 0.05$ and $\delta_{phase} = 0.06$ (note that *x* and *y* axis in Figure 6-6 have different scaling).

*Thres5* is not used in the ODTW, which applies the threshold in each band. In this case, the band onset detection results cannot be compared against the hand labelled database, which provides the onset location information band-wise.

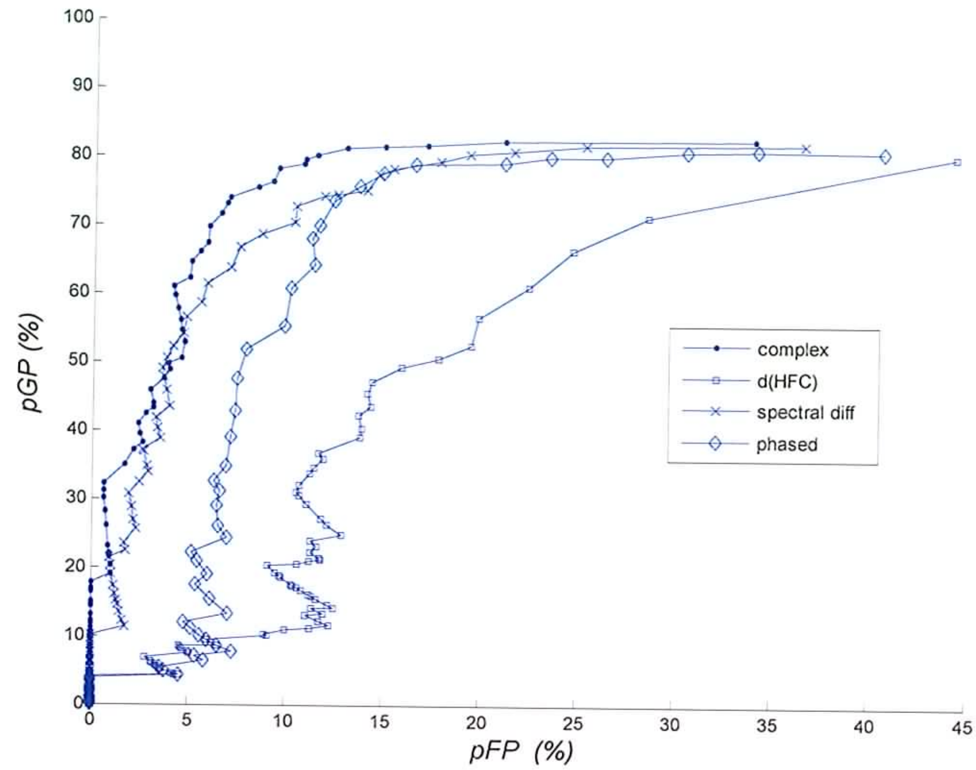

**Figure 6-6: Calculation of the best $\delta$ value for *Thres5* by plotting *pFP* against *pGP*. The methods utilised are shown in the figure legend**

The accuracy results *acc* obtained by evaluating TEST 1 using the existing onset detection methods are shown in Table 6-3.

| | ***Complex*** | | | ***d(HFC)*** | | | ***Spec. diff.*** | | | ***Phase*** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***H / L*** | ***Thres1*** | ***Thres2*** | ***Thres5*** | ***Thres1*** | ***Thres2*** | ***Thres5*** | ***Thres1*** | ***Thres2*** | ***Thres5*** | ***Thres1*** | ***Thres2*** | ***Thres5*** |
| **512/1024** | 57.61 | 42.39 | 68.15 | 15.42 | 22.72 | 41.58 | 44.22 | 38.54 | 62.68 | 5.07 | 16.63 | 62.88 |
| **512/2048** | 47.67 | 37.53 | 66.13 | 17.44 | 22.52 | 40.16 | 49.70 | 37.32 | 64.30 | -2.84 | 31.44 | 59.03 |
| **1024/2048** | 53.55 | 23.94 | 59.23 | 33.87 | 17.24 | 30.43 | 47.67 | 24.14 | 56.19 | 41.78 | 1.01 | 48.28 |

**Table 6-3: Comparison of *acc* results (in %) by evaluating the existing approaches for different *H/L* pairs and the thresholding methods *Thres1*, *Thres2* and *Thres5***

As a comparison, the set of parameters of each onset detection method (existing approaches and the ODTW) that produce the best accuracy results, *acc*, by evaluating TEST 1 are shown in Table 6-4.

| | *acc* | *L* | *H* | *TM* | *pGP (%)* | *pFP (%)* | *pFN (%)* |
|---|---|---|---|---|---|---|---|
| **ODTW** | 68.56 | 2048 | 1024 | *Thres2* | 81.74 | 13.89 | 19.23 |
| **Complex** | 68.15 | 1024 | 512 | *Thres5* | 81.14 | 12.96 | 18.83 |
| ***d*(HFC)** | 41.58 | 1024 | 512 | *Thres5* | 71.20 | 28.52 | 27.73 |
| **Spec. diff.** | 64.30 | 2048 | 512 | *Thres5* | 77.08 | 13.70 | 24.57 |
| **Phase** | 62.88 | 1024 | 512 | *Thres5* | 78.90 | 16.53 | 21.76 |

**Table 6-4: Best *acc* results and best system parameters obtained by evaluating TEST 1 using different onset detection methods**

### TEST 2: Evaluation of *fc* and *win*

As in Section 6.3.1 for the ODTW, the existing approaches are first configured with the best performing set of parameters obtained by evaluating TEST 1, which are shown in the second, third and fourth column of Table 6-4. Then, the onset detection systems are evaluated for the same range of *fc* and *win* as in Section 6.3.1, TEST 2. The best *acc* result obtained by successively evaluating the ODTW and the existing methods for each *win* value of its range is shown in Table 6-5. For each *win* case, *fc* is varied within its entire range. As a comparison, the results provided by the ODTW are also shown.

| *win(ms)* | *ODTW* | *Complex* | *d(HFC)* | *Spectral diff.* | *Phase* |
|---|---|---|---|---|---|
| **20** | 3.45 | 15.01 | -14.60 | 24.14 | 15.82 |
| **25** | 26.98 | 38.54 | 6.90 | 41.18 | 36.51 |
| **30** | 40.57 | 48.88 | 19.88 | 49.90 | 47.67 |
| **35** | 58.22 | 56.80 | 34.48 | 57.20 | 56.80 |
| **40** | 68.15 | 62.27 | 41.78 | 59.84 | 61.26 |
| **45** | 71.60 | 66.73 | 47.87 | 63.29 | 62.47 |
| **50** | 71.81 | 70.18 | 51.12 | 65.31 | 63.89 |

**Table 6-5: best *acc* results obtained by evaluating the onset detection methods *by* varying *win* and *fc***

In Figure 6-7, a visual comparison of the *acc* results obtained by evaluating the existing onset detection methods and the ODTW for the entire range of *fc* is shown. In this case, the *win* value that provides the best results in Table 6-5 is used, which is equal to 50ms.

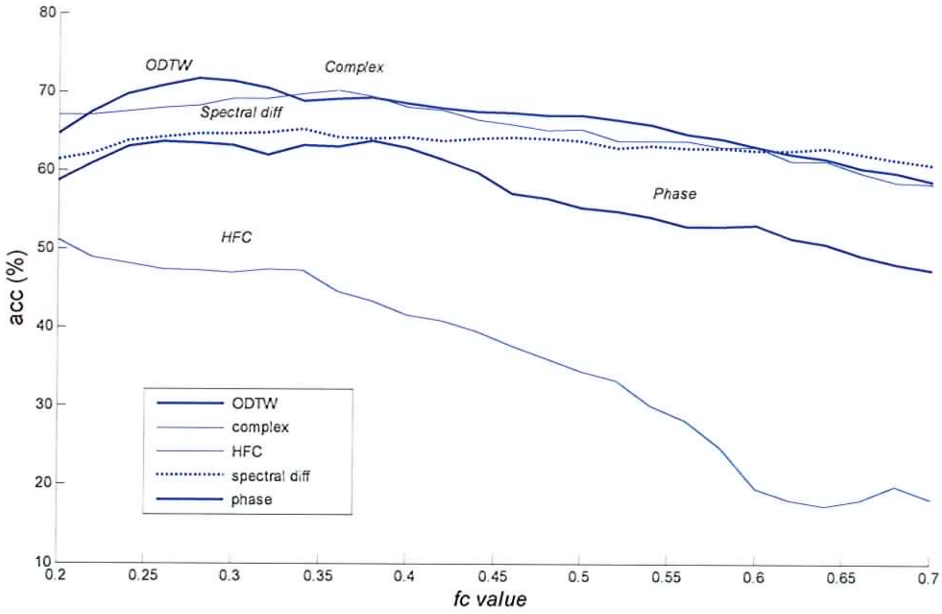

**Figure 6-7: *acc* results (*y* axis) obtained by evaluating the onset detection methods for different *fc* values (*y* axis) and a *win* = 0.05s**

## 6.4 Discussion

In Section 6.3, the performance of the ODTW has been evaluated for a wide range of system parameters and configurations, including the thresholding method *TM*, the pair *H/L* (hop/frame), the LPF cut-off frequency value *fc* and the window length *win*. The results are compared against the performance of existing approaches, which have also been evaluated for their corresponding system parameters and configurations.

The ODTW was first evaluated by using different band thresholding methods and *H/L* pairs. The novel thresholding methods *Thres2* and *Thres4* are compared against the existing *Thres1* method, which provides a very low threshold for the tests performed. This explains why the *pGP* results in Table 6-2 are higher when using *Thres1* than *Thres2* in all cases. The percentage of spurious detections, *pFN*, is also very low. However, *pFP* is much higher than *Thres2*, resulting in a low onset detection accuracy *acc* compared to *Thres2* results.

*Thres4*, which combines *Thres2* and the novel method *Thres3*, provides encouraging results, improving the performance of *Thres1*. However, it was noticed that even though the assumption that the blowing pressure is proportional to the note frequency used in *Thres3* is generally true, the octave jump in the tin whistle has a pressure increase greater than a factor of 2. Thus, the threshold provided by *Thres2* for the reference band, which is usually a high band, produces an excessively high threshold *Thres3* in the low octave bands. This is reflected in the percentage of missed onsets, where *pFN* in Table 6-2 is very high in *Thres4* compared to *Thres1* and *Thres2*.

From Table 6-2, it can be seen that by using the best performing method *Thres2,* the percentage of *pGP* and *pFN* is very similar for the three *H/L* pair cases. The main difference resides in *pFP,* which has a lower value in the *H/L* = 1024/2048 case. This pair produces a smoother energy envelope, which significantly attenuates the effect of the amplitude modulations, as opposed to the use of smaller *L* or *H* lengths that are more sensitive to signal amplitude changes. *pFN* has a low value in all *L/H* pairs, which shows that by using note bands and the thresholding method *Thres2* a low number of events are missed, even in the case of legato playing. However, ornamentation and note events can be separated by less than the *win* length utilised in this test (50ms). In this case, the ODTW only keeps the most prominent onsets, failing to detect both events.

A comparison of the onset detection methods performance (existing approaches and the ODTW) by evaluating TEST 1 is shown in Table 6-4, including the set of parameters of each onset detection method that produce the best *acc* result. Since the tin whistle does not have significant high frequency content, the *d*(HFC) method does not provide accurate results. Ornamentation events produce a quick and rapid energy change before the onset they ornament; the corresponding energy increase can start before the energy of the ornamentation event starts to decrease. In this case, the spectral difference method generates a unique increase in the onset detection function, which commences when the ornamentation event starts increasing and finishes when the energy of the note event stops increasing. By contrast, even though the ODTW also uses energy changes to produce onset peaks in the detection function, the energy of the ornamentation and the note event fall in different bands and will not overlap. The system also improves the phased based approach, which is more sensitive to signal content changes but with the cost of producing more spurious onsets. Finally, Table 6-4 shows that the ODTW and the complex based approach provide similar results.

*Thres2* and *Thres5* are the best performing thresholding methods for the ODTW and the existing approaches respectively. However, it should be noted that, as opposed to *Thres5*, *Thres2* does not require any prior knowledge of the location of the onsets to train the thresholding method. By comparing the onset detection methods using the same thresholding methods *Thres1* and *Thres2*, the best *acc* result is equal to 68.56, which is provided by using *Thres2* (Table 6-2). By contrast, as can be derived from Table 6-3, the best *acc* for all the existing approaches is provided by *Thres1*, whose *acc* value for the complex, HFC, Spectral Difference and Phase based methods is equal to 57.61, 33.87, 49.70 and 41.78 respectively.

The effect that *fc* and *win* parameters have in the ODTW is visually illustrated in Figure 6-5. Since the ODTW utilises the first order difference to obtain the band onset detection function, and uses relatively long frames in the analysis, *win* values higher than 35ms are required to provide accurate results.

In Figure 6-5, it can be seen that at high frequencies, the use of *win* values equal to 45 or 40 ms provide better results than by using a *win* equal to 50 ms. By using low *fc* values the energy envelope is highly smoothed, which can affect the detection of rapid ornamentation events by not reaching the band threshold. By contrast, these ornamentation events are more likely to be detected by using higher *fc* values. In this case, if both ornamentation and note events are separated by a value slightly smaller than

50 ms, *win* values equal to 45 or 40 ms will be capable of detecting both events. Thus, if by decreasing *win* the detection of the remaining onset candidates is not altered, the *acc* value will be increased.

The same evaluation is performed for the existing methods. The best *acc* result obtained per *win* value is shown in Table 6-5, where it can be seen that the ODTW provides better results than the existing methods by using a *win* higher than 30ms. However, its performance considerably degrades for *win* values lower than 35ms.

The ODTW is visually compared against existing methods in Figure 6-7 for different *fc* values and a *win* equal to 50ms. In the figure, it can be seen that approaches that use phase information such as the phase based and the complex based approaches provide better results by using low *fc* values. These approaches are more sensitive to sudden signal changes, and produce noisier onset detection functions that require higher smoothing, which is obtained by using low *fc* values. This phenomenon is more accentuated in the phase based approach. An advantage of the ODTW is that ornamentation and onset events are estimated in different bands. This overcomes the problem that can arise when smoothing an onset detection function in a band-wise configuration. In this case, a note and an ornamentation peak can be merged into a unique peak (note and ornamentation).

## 6.5 Conclusions

A novel onset detector system customised according to the characteristics of the Irish tin whistle has been presented. The system utilises knowledge of the notes and modes that the tin whistle is more likely to produce. The expected blowing pressure that a tin whistle produces per note is also investigated to set the band thresholds *Thres3* and *Thres4*. The development of this novel onset detection approach corresponds to Contribution 1 in Section 1.2.1.

Three novel methods for setting different band thresholds have been introduced, as referred to in Section 1.2.1 as part of Contribution 1. The first method, *Thres2*, produces band thresholds based on the standard deviation. This method provides good results in onset detection functions containing a low level of activity as in the ODTW, improving upon other thresholding methods as [Duxbury '02] (*Thres1*). Another method, *Thres3*, utilises tin whistle musical characteristics to set the threshold value according to expected note blowing pressure. In order to overcome the limitation of requiring a user input to set a band reference threshold, *Thres2* and *Thres3* are combined to form *Thres4*. The results also improve the existing automatic threshold *Thres1*. However, the method accuracy can be affected by the lilt of the player. In order to compensate for the high increase of energy when overblowing to the higher octave, the use of a different band reference threshold in each register could be investigated.

Apart from the different thresholding methods utilised, the ODTW has also been evaluated for other system parameters. The problems related to legato playing are adequately catered for by using a multi-band decomposition, where each band represents one note that the tin whistle can play. The results show that the best system performance is obtained by using the novel thresholding *Thres2*, the pair *H/L* equal to 1024/2048, *win* equal to 50ms and *fc* equal to 0.28*(*fs*/2). However, it has also been shown that smaller *win* values do not require as much smoothing. Such a *H/L* pair value is explained by the occurrence of amplitude modulations within the signal, which is smoothed to a higher extent by using such a pair. This results in a longer delay from the real onset. However, by using such $L$ and a tolerance in the detection established by a *win* higher than 35ms, the percentage of accurate results is high.

By comparing the ODTW against the existing onset detection methods configured with their respective best performing parameters, the ODTW provides the best results. It should be noted that the best results obtained by using the existing methods use a threshold, *Thres5*, that requires the location of the onsets to be configured, which is usually not known. By using the same thresholding methods to compare the onset detection methods, the ODTW improves to a larger extent upon existing methods.

The system presented has been customised for the D key tin whistle, which represents a good example of a slow onset instrument. However, the model is not limited to the tin whistle, and can be configured to other instrument characteristics. As an example, by combining a key detector or an instrument recogniser to the onset detector system, the model can be automatically customised.

### 6.5.1 Limitations of the ODTW

In this chapter, it has been demonstrated that customising the system according to the characteristics of the instrument improves their onset detection accuracy. However, factors inherent to the style of the player can influence the onset detection accuracy. The lilt of the player, which is referred by [Larsen '03] as an element of musical personality that differs between players, gives more stress to some notes than others. This phenomenon will affect the accuracy of *Thres3*.

In addition, it has been noticed that the boundary between C and C# bands is not always clear in the presented implementation; tunes played in modes that do not contain a C# in the structure have more energy in the C# band than in the C band. [Carson '99] documented this phenomenon by stating that the use of C or C# depends on the position of the note in the tune. This statement is corroborated by [Larsen '03], who mentions that when playing C in a quick sequence B-C-D, C# is frequently played instead of C. In the proposed approach, bands were centred according to the standard equal tempered pitch frequencies. However, tin whistle tuning can differ from the equal tempered scale. In addition, different players using the same tin whistle will produce notes "more tuned" than others. These intonation factors depending on the tin whistle and the player will also affect the performance of the method. Thus, a model that customises the system according to the instrument as well as the style of the player is likely to improve the accuracy of the results.

Another limitation resides in the parameter *win*, which in order to reduce the number of spurious detections assume that onset candidates falling within a window length *win* belong to the same onset. However, this also has drawback of missing one event if both ornamentation and note events are separated by less than *win*.

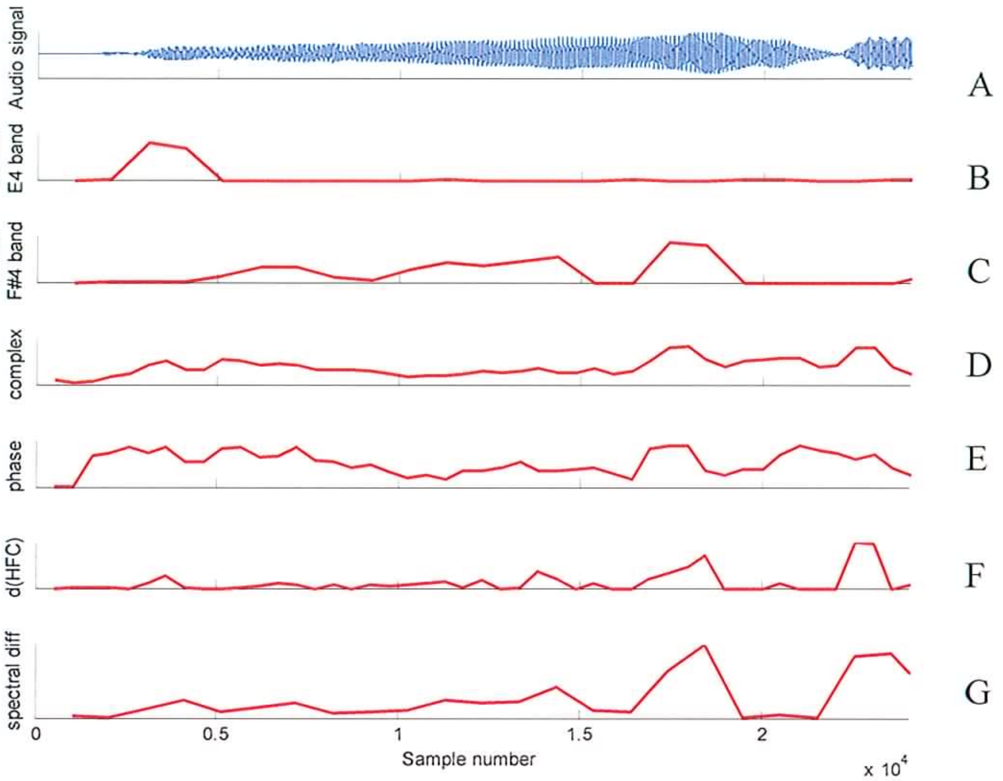

**Figure 6-8: Onset detection functions of the tin whistle signal depicted in plot A by using different techniques, which are labelled in the *y* axis of plots B to G**

Even though the system improves upon existing onset detection methods, it still encounters problems with strong amplitude and frequency modulations. In Figure 6-8 a very slow tin whistle note modulated in amplitude and frequency is depicted in plot A.

The onset detection function of the tin whistle note is obtained by using different onset detection methods. The E4 and F#4 detection bands of the tin whistle based onset detection system are depicted in plots B and C respectively. This method detects the onset in the E4 band 2000 samples after the correct location of the onset. Then, due to the frequency modulation of the note, its energy gradually appears in F#4 band, where a very prominent peak arises due to the amplitude modulation. However, existing methods did not perform any better. This spurious peak is also present in their respective onset detection functions, and the real onset is not clearly discerned in any method.

In the next chapter, the ODTW has been extended to detect single and multi-note ornamentation. The system utilises ornamentation theory introduced in Chapter 2.

# 7 Ornamentation transcription

## 7.1 Introduction

In Chapter 6, an onset detector based on the characteristics of the Irish tin whistle (ODTW), was introduced. It has been shown that the system provides good results on detecting the slow onset of the tin whistle. However, the system considers ornamentation events as onsets, even though ornamentation in Irish traditional music is part of the onset [Larsen '03]. In addition, the system has the limitation of missing one event if both ornamentation and note events are separated by less than a window length defined by the parameter *win* (Section 6.2.3). A method that detects ornamentation and note events separately is presented in this chapter, which transcribes the most commonly played ornamentation types by the tin whistle (Sections 2.2.3 and 2.3) [Gainza '04a]. The system corresponds to Contribution 2 (Section 1.2.1).

A description of the different types of ornamentation played by the Irish tin whistle has been provided in Section 2.3, which also shows that the Irish tin whistle is a good exemplar of the use of ornamentation within Irish traditional music. The ornamentation system presented in this chapter combines the ornamentation knowledge introduced in Section 2.3 with the ODTW presented in Chapter 6.

In Section 7.2, the different parts of the ornamentation transcription are first described. A set of results that evaluate the ornamentation detection system are presented in Section 7.3. Next, a discussion of the results is provided in Section 7.4. Finally, conclusions regarding the ornamentation transcription system are presented in Section 7.5.

## 7.2 System description

The different parts that the ornamentation transcription system are depicted in Figure 7-1. A description of each part is introduced in this section. Firstly, the tasks that the transcription method shares with the ODTW are described, from which vectors of onset and offset candidates are obtained. Then, audio segments are formed and divided into note and ornamentation candidate segments. Next, the system detects single-note ornaments by utilising musical ornamentation theory to establish a set of rules to decide whether a note has been played with single-note ornamentation. Following this, missed ornamentation detections due to strikes played to separate repeating notes are investigated. Finally, multi-note ornaments are formed by combining the estimated single-note ornaments and pitch information.

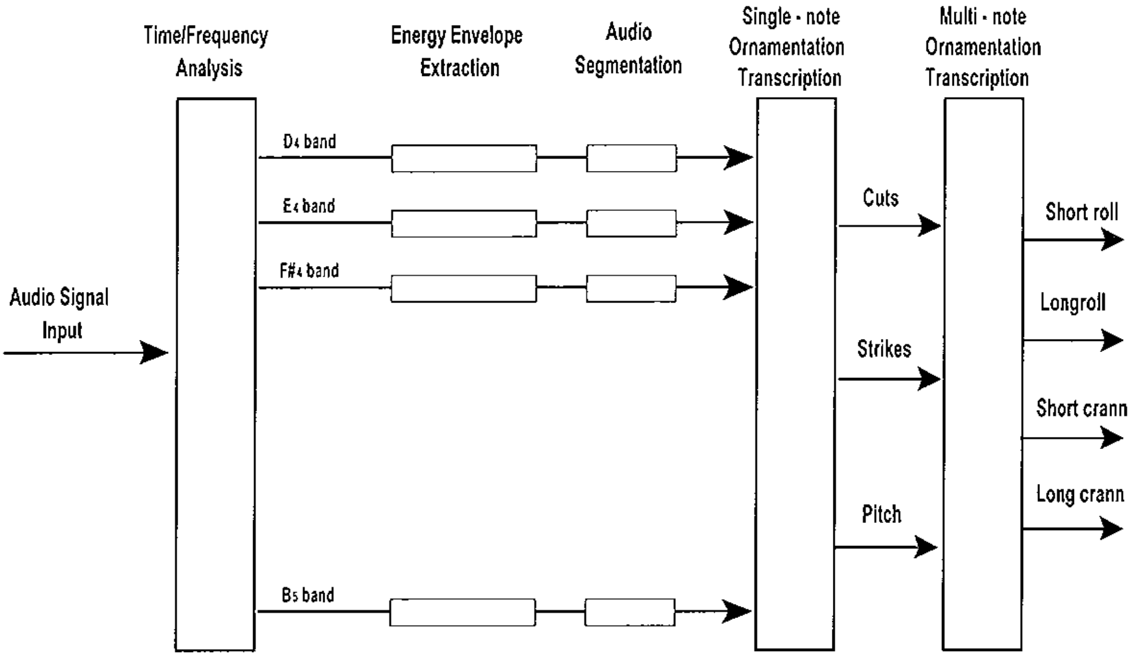

Figure 7-1: Ornamentation transcription system

### 7.2.1 Shared tasks with the onset detection system

As in the ODTW presented in Chapter 6, the signal is first split into 14 overlapping frequency bands. The energy envelope is obtained in each band, and then smoothed using a 3rd order IIR filter. As in Section 6.2.3, the first order difference of the energy envelope is calculated for each band (Equation (84)). Then, the energy increases and decreases are separated into two different vectors, $D_{E(i,n)}$ and $D_{D(i,n)}$ (where $i$ and $n$ are band number and frame number respectively), from which the existence of onset ($t_{on}$) and offset ($t_{off}$) candidate peaks are investigated respectively.

An onset candidate $t_{on}$ is detected if:

$$D_{E(i,n)} = E_{(i,n)} - E_{(i,n-1)} \geq T_i \tag{84}$$

An offset candidate $t_{off}$ is detected if:

$$D_{D(i,n)} = E_{(i,n)} - E_{(i,n-1)} < -T_i \tag{85}$$

where $T_i$ is the band threshold, which is obtained by using the thesholding method *Thres2* introduced in Section 6.2.3. It was shown in Section 6.3 that *Thres2* is the best performing thresholding method for the ODTW, corresponding to part of Contribution 1 (Section 1.2.1).

As an example, Figure 7-2 (plot A) depicts a long roll in $G_4$. Plots B, D and F show the energy envelope of bands F#4, G4 and A4 respectively. In addition, the energy increases and decreases of plots B, D and F are depicted in plots C, E and G respectively. In these plots, energy increases and decreases are depicted with solid and dashed lines respectively.

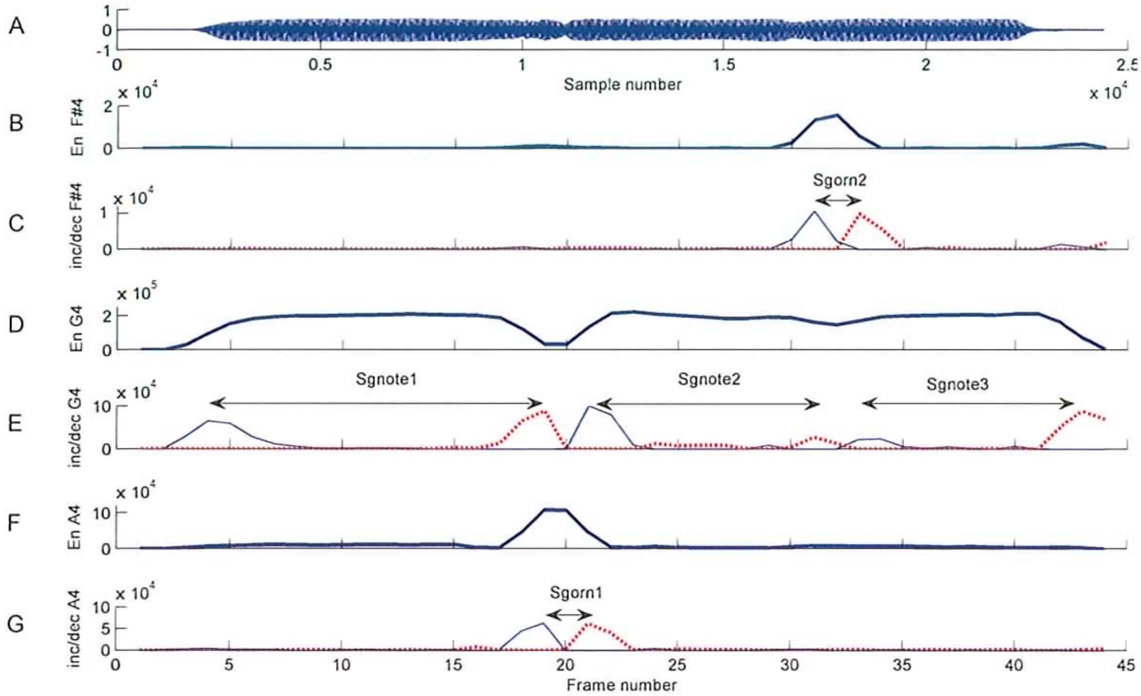

**Figure 7-2: Example of a tin whistle playing a G4 long roll (plot A). Plots B, D and F show the energy envelope of bands F#4, G4 and A4 respectively. The increases and decreases of F#4, G4 and A4 bands are depicted in plots C, E and G respectively.**

### 7.2.2 Audio Segmentation

Every onset candidate $t_{on}$ is matched to the closest offset candidate in time $t_{off}$ (where $t_{off} > t_{on}$) to form audio segments $Sg = [t_{on}, t_{off}]$.

Next, according to time duration, the audio segments are split into note and ornamentation segments as follows:

$$Sg = Sg_{orn} \qquad \text{if } t_{off} - t_{on} < T_e \tag{86}$$

$$Sg = Sg_{note} \qquad \text{if } t_{off} - t_{on} > T_e \tag{87}$$

where $T_e$ is the longest expected ornamentation time for an experienced player.

The duration of a cut or strike provides a measure of how well the ornamentation has been played. Beginners play long cuts or strikes that sound as individual notes [Larsen '03], which will be estimated as $Sg_{note}$ in the system presented. By contrast, experienced players will produce short cuts or strikes below the threshold $T_e$. In this case, the ornaments will embellish the associated notes without sounding as individual notes.

From Figure 7-2, plot E, by appropriately thresholding the band detection function, three note segments will be formed in the band that corresponds to $G_4$ (band $i = 4$):

- $Sg_{note1} = [t_{on}(Sg_{note1}), t_{off}(Sg_{note1})] = [D_{E(4,4)}, D_{D(4,19)}]$
- $Sg_{note2} = [D_{E(4,21)}, D_{D(4,31)}]$
- $Sg_{note3} = [D_{E(4,33)}, D_{D(4,43)}]$

In addition, two ornamentation segments will be formed in bands F#$_4$ ($i$ =3) and A$_4$ ($i$=5):

- $Sg_{orn1} = [t_{on}(Sg_{orn}), t_{off}(Sg_{orn})] = [D_{E(5,19)}, D_{D(5,21)}]$
- $Sg_{orn2} = [D_{E(3,31)}, D_{D(3,33)}]$

Note and ornamentation band segments are separately combined, and sorted in $t_{on}(Sg_{note})$ and $t_{on}(Sg_{orn})$ time order respectively. Finally, a sliding window *win* centred at the $t_{on}$ of each segment candidate is applied, and the segment with the most prominent onset value ($t_{on}$) is maintained as a segment candidate. This operation is performed separately for the ornamentation and note segments.

In order to cancel spurious segment detections caused by amplitude modulations, the following method is applied. If a band amplitude envelope produces two consecutive energy increases, $t_{on}(Sg_{note1})$ and $t_{on}(Sg_{note2})$, associated to the same offset candidate

($t_{off}(Sg_{note1})$ = $t_{off}(Sg_{note2})$), it is assumed that the energy increases have been due to an onset and an amplitude modulation respectively. Thus, the note segment whose onset is delayed relative to the other segment onset is discarded. As an example, if $t_{on}(Sg_{note1}) > t_{on}(Sg_{note2})$, $Sg_{note1}$ is rejected.

### 7.2.3 Single-note ornamentation Transcription

As can be seen in Figure 7-1, once the detected segments have been split into note and ornamentation segments, $Sg_{note}$ and $Sg_{orn}$, the existing single-note ornaments are transcribed. To decide whether a note represented in $Sg_{note}$ has been played with the ornamentation represented in $Sg_{orn}$, ornamentation theory is applied. The different rules comprising the single-note ornamentation task are depicted in Figure 7-3. First, the main ornamentation rule used to associate ornamentation and note segments is defined. Then, rules to detect cut and strike ornaments are introduced, including some exceptions where the ornamentation detection rules do not apply. Finally the single-note ornaments are transcribed.

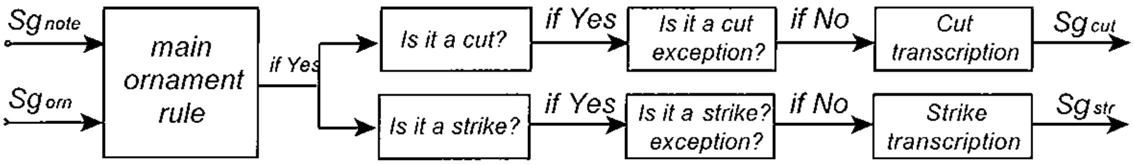

**Figure 7-3: Single-note ornamentation transcription block diagram**

**Main ornamentation rule**

In Irish traditional music, single-note ornamentation is played right on the beat, providing an accurate time for start of a new note [Larsen '03]. Thus, a note segment occurs soon after the ornamentation segment:

$$|t_{on}(Sg_{note}) - t_{off}(Sg_{orn})| < T_l \qquad (88)$$

where the threshold $T_l$ ensures that the segments are connected in legato. The absolute value copes with the case where the instrument offset has a slower profile than its onset. In this case, $t_{off}(Sg_{orn})$ can be delayed from $t_{on}(Sg_{note})$.

An example of the application of the rule is illustrated in Figure 7-2, where the ornamentation segments $Sg_{orn1}$ and $Sg_{orn2}$ will be associated to the note segments $Sg_{note2}$ and $S_{note3}$ respectively.

**Cut detection rules**

- *Main cut detection rule*: As previously shown in Section 2.3, the cut momentarily increases the pitch. Thus, if a cut has been played, then $i(Sg_{orn}) > i(Sg_{note})$, where $i$ denotes the band number. From Table 2-1, bands A4 and G4 have value $i$=5 and $i$=4 respectively. Consequently, $i(Sg_{orn1})$ and $i(Sg_{note2})$ have also value $i$=5 and $i$=4 respectively, which follows the rule $i(Sg_{orn1}) > i(Sg_{note2})$.
- *Exception: cuts in descending notes separated by more than a 2^nd^ musical interval.* The tin whistle is commonly played in legato, which means that successive notes are connected without any intervening silence. When descending in the melody between two notes separated by more than a 2nd

within the same octave, the uninterrupted flow of air can produce energy in the bands located between the two successive notes in the melody. Since this only occurs for a short period of time, ornamentation segment candidates can arise in between the bands of the two note segments. If the estimated ornamentation segment is located in a higher band than the second note segment and the segments follow the main ornamentation rule, the ornamentation segment will be wrongly detected as a cut. However, by analysing the physical mechanism of cutting a note using different fingering techniques, spurious detections of cuts can be corrected.

The fingering techniques considered are the standard fingering techniques [Brother '06, Larsen '03], and a new fingering technique proposed by [Larsen '03], which is being adopted by many Irish traditional musicians:

1. *Standard fingering techniques:*

By using standard fingering techniques, cutting a note comprised in the [D –G] interval of Table 6-1 is performed in the G hole [Brother '06, Larsen '03] (located at the third hole in order from the embouchure), which produces an ornamentation in band A. Thus, if the melody descends from a note higher than A and cuts a note comprised in the [D –G] interval, an ornamentation segment will arise in band A, which is located in between both melody note bands. In all the remaining cases, an ornamentation segment cannot arise in between the note bands. It should be noted that this rule only applies when both notes are in the same register.

2. *Larsen's fingering techniques:*

By using Larsen's fingering techniques, the holes that are in between the interval cannot be utilised to produce the cut [Larsen '03]. These holes are required to produce the new note and cannot be used for cutting at the same time. As a result of this, different fingering is required to produce the cut, this is accomplished by momentarily uncovering the lowest covered hole that produces the first note of the melody [Larsen '03]. For example, moving from A4 to D4 and cutting with D4, A4 hole (located at the second hole in order from the embouchure) is uncovered. This produces a segment in B4 band.

Since the style of the player is not known by the ornamentation detector, only rules shared by both fingering techniques will be utilised. Thus, if an ornamentation segment fulfils the following conditions:

- The segment is located in between descending notes separated by more than a 2nd
- Both notes are within the same octave
- The first note is lower than B

it is deemed that the segment has not been generated by the action of cutting a note.

**Strike detection rules:**

- *Main strike detection rule*: the strike lowers the pitch. As opposed to the cut, where fingering varies depending on the player style, the strike is always played by covering the closest available hole [Larsen '03]. Thus, if a strike is played, then $i(Sg_{orn}) = i(Sg_{note}) - 1$. From Table 2-1, bands F#4 and G4 are equal to 3 and 4 respectively. Consequently, in Figure 7-2 $i(Sg_{orn2})$ and $i(Sg_{note3})$ are also equal to 3 and 4 respectively, following the rule $i(Sg_{orn2}) = i(Sg_{note3}) - 1$.
- *Exception 1*: Strikes that occur when the melody ascends between notes that are in the same register cannot be played [Larsen '03].
- *Exception 2*: If the "ascending" strike breaks through the second register, the ascending interval cannot be greater than an octave [Larsen '03].

**Ornamentation transcription and improvement of the onset time accuracy**

If an ornamentation and a note segment, $Sg_{orn}$ and $Sg_{note}$, are detected and follow the set of rules defined above, it will be considered that the note represented in $Sg_{note}$ is played with the ornamentation represented in $Sg_{orn}$. As stated earlier, ornamentation in Irish traditional music is played right on the beat, providing an accurate time of new note commencement [Larsen '03]. Thus, the system is also utilised to improve the accuracy of the onset estimation, since an ornamented note will be considered as just one note with its onset starting when the ornamentation commences. Consequently, the ornamented note segment will comprise of: $[t_{on}(Sg_{orn}), t_{off}(Sg_{note})]$, and it will be denoted as $Sg_{cut}$ or $Sg_{str}$, if $Sg_{orn}$ corresponds to either a cut or a strike respectively.

From Figure 7-2, it can be seen that two modified segments will be formed, which are given by:

- $Sg_{cut}= [D_{E(5.19)}, D_{D(4,31)}]$
- $\mathrm{Sg}_{\mathrm{str}}= [D_{E(3,31)}, D_{D(4,43)}]$.

### 7.2.4 Correction of missed strike detections

Having estimated the single-note ornaments cuts and strikes, segment estimation errors due to missed strikes are next investigated. The most common use of the strike is to separate repeating notes [Duggan '06b, Larsen '03], as it occurs between the first and second note of a short and long roll, the second and third note of a long roll, and in cranns (Section 2.3). Based on the set of rules described in Section 7.2.3, when this type of strike is played an ornamentation segment should arise one band below the two note segments. In addition, the ornamentation segment should be located in between the two band note segments. Physically, the only movement of the players fingers is to rapidly cover the first uncovered hole without interrupting the flow of air [Larsen '03]. Due to the brevity of the strike, the ornament can be missed by the ornamentation transcription system. Three commonly occurring scenarios, wherein a strike that separates two notes (*note1* and *note2*) is not correctly detected have been identified:

- **Scenario 1:** the onset and offset peaks that form the ornament segment are not prominent enough to reach the band threshold. Consequently, the ornamentation segment is not detected.
- **Scenario 2:** The division between the two note segments $Sg_{note1}$ and $Sg_{note2}$ is not clear. As a result of this, the offset of the first repeating note $t_{off}(Sg_{note1})$, and

the onset of the second repeating note $t_{on}(Sg_{note2})$ are missed. In this case, both notes are estimated as a unique note segment: $[t_{on}(Sg_{note1})$ , $t_{off}(Sg_{note2})]$.

- **Scenario 3:** both Scenario 1 and Scenario 2 occur. The ornamentation segment is missed, and both notes are estimated as a unique note.

In an effort to detect the missed strikes in the above introduced scenarios, the following method is applied:

*1.* First, a more relaxed threshold $T'i$ is applied in each band to estimate new ornamentation segment candidates, which are denoted as $Sg_{orn'}$.

*2.* Next, the occurrence of the three scenarios mentioned above is investigated:

- **Is Scenario 1 occurring?** If a $Sg_{orn'}$ is located one band below the two note segments $Sg_{note1}$ and $Sg_{note2}$, and $|t_{on}(Sg_{orn'})$ - $t_{off}(Sg_{note1})| < T_l$, and $|t_{on}(Sg_{note2})$ - $t_{off}(Sg_{orn'})| < T_l$, it will be considered that Scenario 1 has occurred. In this case, a new ornamentation segment is formed, where $Sg_{str} = [t_{on}(Sg_{orn'}), t_{off}(Sg_{note2})]$.
- **Is Scenario 2 occurring?** If a $Sg_{orn}$ arises at the middle of a note segment $Sg_{note}$, and it is located one band above the ornamentation segment, it will be considered that scenario 2 has occurred. In this case, $Sg_{note}$ will be split into two different segments: $Sg_{note1}$ and $Sg_{str}$, where $Sg_{note1} = [t_{on}(Sg_{note}), t_{on}(Sg_{orn})]$ and the modified note segment $Sg_{str} = [t_{on}(Sg_{orn}), t_{off}(Sg_{note})]$.
- **Is Scenario 3 occurring?** If the ornamentation segment candidate in the previous scenario 2 has been estimated using $T'i$ instead of $Ti$, which also signifies that $Sg_{orn'}$ has been detected as opposed to $Sg_{orn}$. Then, Scenario 3 arises.

In Figure 7-2, a strike has been played to separate two notes (*note2* and *note3*), where it can be seen that $t_{off}(Sg_{note2})$ and $t_{on}(Sg_{note3})$ have less prominent peaks. If the band threshold $T_4$ is set above the peak values, $Sg_{note2}$ and $Sg_{note3}$ will be estimated as one unique segment $Sg_{note}$. In this case, Scenario 2 will occur and $Sg_{note}$ will be split into two separate segments $Sg_{note2}$ and $Sg_{str}$.

### 7.2.5 Multi-note ornamentation Transcription

As introduced in Section 2.3, multi-note ornamentation can be seen as the process of playing certain combinations of unornamented and ornamented slurred notes having the same pitch. Thus, by estimating both note pitch and single-note ornamentations, multi-note ornamentation can be transcribed.

In Section 7.2.3, a method that transcribes single-note ornamentation has been introduced. The method can also be interpreted as a pitch detector, since the band where a new note segment arises corresponds to the pitch of the note. Thus, if a note segment is detected in band $i$, $i(Sg_{note})$, the position of $i$ in Table 2-1 will correspond to the note pitch. As for example, if $i(Sg_{note})=1$, the note played is D4. In the case of a modified segment, the note pitch $i(Sg_{note})$ that forms the modified segment will correspond to the band of the note segment $i(Sg_{cut})$ or $i(Sg_{str})$.

Applying the theory introduced in Section 2.3, the following combinations of segments will form multi-note ornamentations:

- **Long roll:** $Sg_{lr} = [Sg_{note}, Sg_{cut}, Sg_{str}]$, where:
  - $i(Sg_{note}) = i(Sg_{cut}) = i(\mathrm{Sg}_{\mathrm{str}})$

  - $|t_{on}(Sg_{cut}) - t_{off}(Sg_{note})| < T_l$
  - $|t_{on}(Sg_{str}) - t_{off}(Sg_{cut})| < T_l$
- **Short roll:** $Sg_{sr} = [Sg_{cut}, Sg_{str}]$, where:
  - $i(Sg_{cut}) = i(Sg_{str})$
  - $|\mathrm{t_{on}}(Sg_{str}) - \mathrm{t_{off}}(Sg_{cut})| < T_l$
- **Long crann:** $Sg_{lc} = [Sg_{note}, Sg_{cut1}, Sg_{cut2}]$, where:
  - $i(Sg_{note}) = i(Sg_{cut1}) = i(Sg_{cut2})$
  - $|t_{on}(Sg_{cut1}) - t_{oft}(Sg_{note})| < T_l$
  - $|t_{on}(Sg_{cut2}) - t_{off}(Sg_{cut}1)| < T_l$
- **Short crann:** $Sg_{sc} = [Sg_{cut1}, Sg_{cut2}]$, where:
  - $i(Sg_{cut1}) = i(Sg_{cut2})$
  - $|t_{on}(Sg_{cut2}) - t_{off}(Sg_{cut1})| < T_l$

By way of an example, a long roll will be detected in the example depicted in Figure 7-2 as $Sg_{lr} = [Sg_{note1}, Sg_{cut}, Sg_{str}]$, where:

- $Sg_{note1} = [D_E(4,4), D_D(4,19)]$
- $Sg_{cut} = [D_E(5,19), D_D(4,31)]$
- $Sg_{str} = [D_E(3,31), D_D(4,43)]$

## 7.3 Ornamentation Results

The same database of tin whistle signals utilised in Chapter 6 for the ODTW is used to evaluate the performance of the ornamentation transcription system presented here. The performers utilised the whole range of ornamentation types considered by the ornamentation detection system: cuts, strikes, rolls and cranns. The presented system developed is more complex than an onset detector, since an incorrect onset or offset detection in a note or ornamentation segment, would result in incorrect ornamentation transcription. In addition, an incorrect single-note ornamentation segment detection or an incorrect pitch detection of the notes that comprise the multi-note ornament, will result in an incorrect multi-note ornamentation transcription. In addition, the ornament database is smaller than the actual notes played, which increases the difficulty of the testing procedure.

The thresholding method based on the standard deviation has again been utilised, as introduced in Section 6.2.3 (denoted as *Thres2*). As shown in Section 6.3, this method provides the best results in the ODTW.

Firstly, the performance of the algorithm is evaluated on detecting single-note ornaments for the following parameters, which description was provided in Section 6.3.1:

- *L/H*: the pair of system parameters *L/H* equal to 512/1024, 1024/2048 and 512/2048 are considered.
- ***win:*** different values scaled by steps of 5ms within the [20:50ms] range.
- *fc*: different values scaled by steps of 0.02*(*fs/2*) within the [0.2:0.7]*(*fs/2*) Hz range.

Since multi-note ornamentation is formed by combining single-note ornamentation, the accuracy of the multi-note ornamentation detector depends on the performance of the single-note ornamentation detector. Consequently, instead of evaluating the multi-note ornamentation detector for the whole range of *fc* and *win* parameters, the *fc* and *win* values that provide the best results in Section 7.3.1 are utilised to detect the multi-note ornaments. The results obtained are shown in Section 7.3.2.

### 7.3.1 Single-note ornamentation results

The system is first evaluated on detecting single-note ornamentation for the different system parameters. In Chapter 6, the number of *FN* and *FP* was separately obtained to estimate the onset detection accuracy. By using the same method, an ornamentation whose location has been correctly detected, but whose type has been wrongly transcribed, (e.g., a strike detected as a cut at the correct location) will increase both *FN* and *FP*. This will doubly affect the *acc* result by application of Equation (83). Consequently, in order to measure the accuracy of the detection, the *acc* equation has been simplified. If a single-note ornament fulfils the following conditions, the detection has been correctly detected, denoted as *Corr*:

- *Single-note ornamentation location:* the $t_{on}(Sg_{orn})$ of the detected single-note ornament $Sg_{cut}$ or $Sg_{str}$, should fall within a window length *win* centred at the commencement time of the target ornamentation.
- *Single-note ornamentation type*: the type of detected single-note ornament $Sg_{cut}$ or $Sg_{str}$ should coincide with the target single-note ornamentation type.

Then, *pCorr* is calculated by dividing the number of *Corr* by the total of ornamentation events in the database. If one or both of the above listed conditions is not correct, the detection is rejected as false, denoted as *Fal*. Finally, the *acc* measure is now given by:

$$acc = \frac{total - Fal}{total} * 100\% \qquad (89)$$

In Table 7-1, the best *acc* results obtained per *H/L* pair for the entire range of *fc* and *win* are first shown, where the *fc* that provides the best results varies depending on the pair utilised (*SN Orn* in Table 7-1 denotes Single-Note Ornamentation).

| *L / H* | *Notes* | *SN Orn* | *fc* | *win* | *pCorr(%)* | *acc (%)* |
|---|---|---|---|---|---|---|
| 512 / 1024 | 493 | 85 | 0.28 | 50ms | 62.35 | 56.47 |
| 512 / 2048 | 493 | 85 | 0.30 | 50ms | 62.35 | 55.29 |
| 1024 / 2048 | 493 | 85 | 0.64 | 50ms | 65.88 | 51.76 |

**Table 7-1: Single-note ornamentation results using different *L/H* pairs**

<u>Evaluation of the parameter *win*</u>

In order to investigate the impact of the parameter *win* on the detection accuracy, the best *acc* result obtained by successively evaluating the system for each *win* value of its range is shown in Table 7-2. For each *win* case, *fc* is varied within its entire range. From Table 7-2, the best performing *L/H* pair in Table 7-2 is equal to 512/1024 for all *win* values.

| | | *win* | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | **20** | **25** | **30** | **35** | **40** | **45** | **50** |
| ***L/H*** | **512/1024** | 29.41 | 49.41 | 51.76 | 52.94 | 52.94 | 55.29 | 56.47 |
| | **512/2048** | 28.24 | 44.71 | 44.71 | 49.41 | 51.76 | 54.12 | 55.29 |
| | **1024/2048** | 2.35 | 21.18 | 24.71 | 41.18 | 50.59 | 51.76 | 51.76 |

**Table 7-2: *acc si*ngle-note ornamentation results by varying *win*.**

### 7.3.2 Multi-note ornamentation results

As introduced in Section 7.2.5, multi-note ornaments are formed by combining single-note ornamentation and pitch information. Thus, a correct multi-note ornament should fulfil the following conditions:

- *Multi-note ornament location:* the $t_{on}(Sg_{orn})$ of the detected multi-note ornament $Sg_{lr}$ . $Sg_{sr}$, $Sg_{lc}$ or $Sg_{sc}$ should fall within a window length 2**win,* which is centred at the time when the first segment of the target multi-ornament commences (e.g: $t_{on}(Sg_{note})$ in a $Sg_{lr}$).
- *Multi-note ornamentation type*: The type of detected multi-note ornament $Sg_{lr}$ . $Sg_{sr}$, $Sg_{lc}$ or $Sg_{sc}$ is the same as the target multi-note ornamentation type.

It should be noted that a more tolerant *win* has been used in this case, *2*win* as opposed to *win*. If the type of detected and target multi-note ornaments are the same, it signifies that the various segments that comprise the detected multi-note ornament have been correctly identified. Consequently, a separation of 2**win* between the target and candidate multi-note ornament is an accurate estimation of a correct multi-note ornament.

The system utilises the *fc* and *win* that provide the best *acc* results on detecting single-note ornaments, whose values are derived from Table 7-1. In Table 7-3, the *acc* results obtained on detecting multi-note ornaments by using these parameters are shown (*MN Orn* in Table 7-1 denotes Multi-Note Ornamentation).

| *L / H* | *Notes* | *S N Orn* | *MN Orn* | *pCorr(%)* | *acc (%)* |
|---|---|---|---|---|---|
| **512 / 1024** | 493 | 85 | 21 | 42.86 | 42.86 |
| **512 / 2048** | 493 | 85 | 21 | 42.86 | 42.86 |
| **1024 / 2048** | 493 | 85 | 21 | 47.62 | 47.62 |

**Table 7-3: Multi-note ornamentation results**

## 7.4 Discussion

In Section 7.3, the performance of the ornamentation detector has been evaluated on detecting single and multi-note ornaments. Firstly, the method has been evaluated for the single-note ornamentation case by using different *H/L* pairs, LPF cut-off frequency *fc* and window length *win* values. The results obtained are shown in Table 7-1. It can be seen that the best results arise for an *H/L* pair of 512/1024, as opposed to the ODTW in which an *H/L* of 1024/2048 is the best option (Table 6-2). In order to form a segment, both onset and offset times have to be correctly estimated, from which the offset generally has a slower profile. This causes problems by using an *H/L* pair equal to 1024/2048, which smoothes the ornamentation event in the energy envelope more than the other pairs. Due to its long duration, the offset is detected with a greater delay to its real time location than the onset from its corresponding real location. This has the result of lengthening the segment, which can be detected as a note segment as opposed to an ornamentation segment. In addition, the offset of the segment cannot be prominent enough to reach the band threshold, which results in a missed segment detection. From Table 7-1, the best *acc* result by using an *H/L* pair equal to 1024/2048 is obtained for a *fc* value (0.64*(*fs*/2)). The use of low *fc* values will degrade the results using an *L/H* = 1024/2048. Such smoothing produces slower offsets, which has the result of accentuating the aforementioned problems that such pairs encounter in the analysis. In contrast, the other pairs provide the best results when *fc* is set to a lower value (approximately equal to 0.3*(*fs*/2)).

In the ornamentation detection system, detected segments are combined only with segments of the same type (ornamentation or note segments). Thus, the parameter *win* only sets the tolerance in the detection accuracy. This is reflected in Table 7-2, where the *acc* estimations improve with the *win* value for all pairs. Less accurate results are provided by the use of the *L/H* pair equal to 1024/2048, whose accuracy degrades for values smaller than 35 ms.

It should be noted that the database of tin whistle signals also comprise slides and vibratos, which are not detected by the system. However, these strongly frequency modulated ornaments do not affect the accuracy of the detection of the remaining ornaments, since their characteristics do not match with the rules entered in the system.

The system has also been evaluated on detecting multi-note ornaments. The difficulty of detecting these ornaments must be emphasised. As an example, a correct detection of a long roll implies correctly estimating three note segments ($Sg_{note1}$, $Sg_{note2}$ and $Sg_{note3}$), two ornamentation segments ($Sg_{orn2}$ and $Sg_{orn3}$), three note pitches ($i(Sg_{note1})$, $i(Sg_{note2})$ and $i(Sg_{note3})$), and also the bands where the two occurring ornamentation occurring band segment ($i(Sg_{orn2})$ and $i(Sg_{orn3})$) arise. As can be seen in Table 7-3, the percentage of correct detections, *pCorr* has the same value as the *acc* result for all the pairs. This is explained by the criteria utilised by the system, where correct estimations have to fulfil a large number of conditions. This provides the result of not detecting any spurious multi-note ornaments. Consequently, *Fal* in the *acc* equation only contains missed detections. In this case, *acc* and *pCorr* provide the same result.

## 7.5 Conclusions

There are many different styles of playing ornamentation within Irish Traditional music. Consequently, such an improvised element of musical expression cannot be fully defined by a set of rules. However, the most common types of ornaments are transcribed by the novel ornamentation detection system presented in this chapter, which represents the first step towards a fully transcription of ornamentation within Irish Traditional music. The system combines the ODTW presented in Chapter 6 with the ornamentation theory introduced in Section 2.3, and which is entered into the system by defining a set of rules. The criteria include single and multi-note formation rules, exceptions where the rules do not apply and typical scenarios where strike detections can be missed. The development of the system corresponds to Contribution 2 in Section 1.2.1.

The percentage of correct single and multi-note ornamentation detections is relatively high in both cases. Since the system inherits the main structure from the ODTW, the majority of the conclusions and limitations of the ODTW also apply to the ornamentation system (Section 6.5). However, the system overcomes a number of those limitations: in the ODTW, the band onset candidate peaks are combined, and the strongest peak within a window *win* is selected as the unique onset candidate. Thus, if both ornamentation and onset event peaks fall within the same window, one event is missed. This is not the case in respect of the ornamentation detector, since the segments detected are combined only with segments of the same type (ornamentation or note segments). In addition, the system provides a better time estimation, since the onset is estimated at the beginning of the ornamentation event $t_{on}(Sg_{orn})$. This onset time estimation better reflects the characteristics of ornamentation within Irish Traditional music, wherein the ornamentation is considered as part of the onset.

The method presented in this chapter limits the transcription to cuts, strikes, rolls and cranns. Transcribing more types of ornamentation might be considered as an area of future work. Thus, the creation of a corpus of different styles of playing ornamentation by different tin whistle players should be undertaken. This will permit analysing the technical characteristics of each ornamentation type within each different style of playing.

In the following chapter, a second onset detector is presented, with implementation based on the comb filter techniques introduced in Chapter 3.

# 8 Onset Detection system based on Comb Filters (ODCF)

## 8.1 Introduction

In chapter 4, a literature review of onset detection was presented. It was documented that existing onset detection methods encounter difficulties in dealing with amplitude and frequency modulations, during fast passages such as legatos and ornamentations, and also with very slow onsets. In Chapter 6, an onset detection system based on the tin whistle characteristics, ODTW, was demonstrated. The results show that by adequately thresholding the different frequency bands, the ODTW could partially deal with the difficulties associated with amplitude modulations. The system utilises a multi-band decomposition, with one band per note, which reduces the problems related to legato playing. In order to detect ornamentation, an extension of the ODTW was presented in Chapter 7. The ODTW utilises the first order difference to calculate the band onset detection functions, which produces a delay from the correct onset time. Nevertheless, by using a frame length $L$ =2048 and a tolerance window $win$ larger that 35 ms, the onsets are accurately estimated. However, it has been shown that the system is still vulnerable to strong amplitude modulations and frequency modulations. As discussed in Section 4.5, this is also the case for the existing onset detection approaches, where energy and phase based approaches are prone to detect spurious onsets when dealing with amplitude and frequency modulations respectively.

In order to deal with these problems, a novel onset detector has been implemented. Existing onset detectors utilise energy and/or phase information to generate an onset detection function. In contrast, the onset detection system presented in this chapter utilises the harmonicity changes of the signal by using comb filters, which also have a harmonic type of magnitude response. In addition to dealing with signal modulations, this method provides a more accurate onset time. This novel onset detection system represents Contribution 3 of this thesis [Gainza '05b]. Comb filter techniques are also utilised in Chapter 9 to construct a pitch detector.

In this chapter, the different parts of the ODCF are first introduced in Section 8.2. Next, a wide range of tests have been performed in Section 8.3 to validate the approach. Finally a discussion of the results is given in Section 8.4, which leads to the conclusions documented in Section 8.5.

## 8.2 Onset detection system

In this chapter, a technique for detecting note onsets using FIR comb filters which have different filter delays is presented [Gainza '05b]. The onset detector focuses on the harmonic characteristics of the signal, which are calculated relative to the energy of the frame. Both properties are combined by utilising FIR comb filters on a frame by frame basis. In order to generate an onset detection function, the changes of the signal harmonicity are tracked. This produces peaks in the harmonicity changes that a new onset provides in the signal.

The method relates the harmonicity detection to the energy of the analysing frame, which is suitable for detecting slow onsets, and provides an accurate onset estimation time. The approach is robust for dealing with amplitude modulations; if the energy of the signal changes between successive frames (but not its harmonicity) the onset detection function remains stable. In addition, the method is robust to frequency modulations that gradually occur in the signal, since the signal harmonicity does not change considerably between frames.

In Figure 8-1, a block diagram illustrating the different components of the system is depicted.

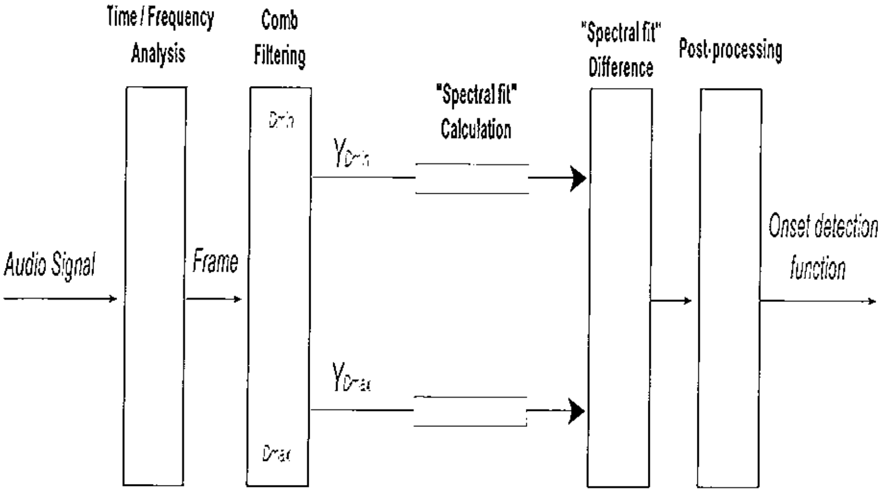

**Figure 8-1: Onset Detection system based on comb filters (ODCF)**

This section describes the different blocks of the ODCF. A time - frequency analysis is first required. Then, by using the comb filter techniques introduced in Chapter 3, different filter outputs are obtained. Next, a measure denoted as "spectral fit" is calculated, which is utilised to obtain the onset detection function. Finally, some post-processing tasks are applied.

### 8.2.1 Time-Frequency Analysis and filtering

As in Section 6.2.1, the frequency evolution over time is obtained using the Short Time Fourier Transform (STFT), which is calculated using a Hanning window. The frame representation in the frequency domain $X(m,k)$, where $m$ and $k$ are the frame and bin numbers respectively, is fed into a bank of FIR comb filters, which are implemented using the filter techniques introduced in Chapter 3.1. All the FIR comb filters of the bank are built using a value $g=1$. However, each filter $i$ uses a different delay $D_i$ from a comb filter bank delay vector $D = [D_{min}...D_{max}]$, where $D_{min}$ and $D_{max}$ are the shortest and longest delays of the vector. Next, the filter output $Y_{Di}(m,k)$ is calculated as follows:

$$Y_{D_i}(m,k) = X(m,k) \times H(D_i,k) \tag{90}$$

where $H(D_i,k)$ denotes an FIR comb filter frequency response built with a delay $D_i$.

As an example, Figure 8-2 depicts the magnitude response of four FIR comb filters, with delays corresponding to the periods of the first four semitones of octave 4.

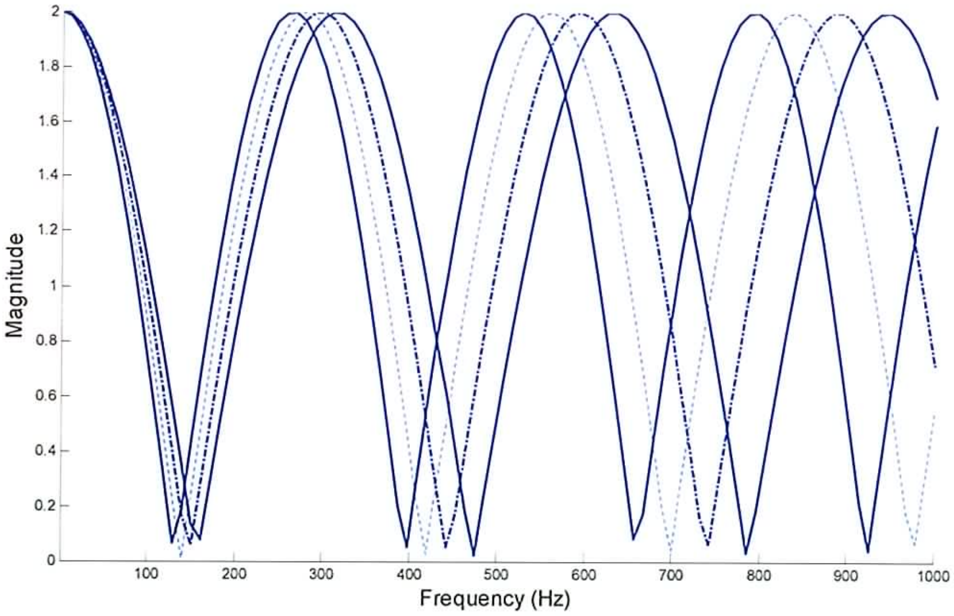

**Figure 8-2: Magnitude response of four comb filters, whose delays *D* correspond to the period in samples of the four semitones of octave 4.**

### 8.2.2 Spectral fit calculation

The energy of each output is calculated in the frequency domain as follows:

$$E(m,D_i) = \sum_{k_i=1}^{M} \left\{ \left| Y_{D_i}(m,k)^2 \right| \right\} \tag{91}$$

where $M$ denotes the FFT length.

From Equation (3), it can be seen that the maximum output amplitude that the FIR comb filters can reach with $g = 1$ is $y_{max}(n) = 2*x(n)$, which can only occur for the case of $x(n) = x(n+D)$. In this case, the maximum output energy is $E(y_{max}) = 4*x^2(n)$. Then, by normalising each output energy $E(m,D_i)$ with $E(y_{max})$, a measure of how similar the filter $H(D_i,k)$ is to the perfect FIR comb filter that extracts the maximum energy $E(y_{max})$ is obtained:

$$E_m(m,D_i) = \frac{E(m,D_i)}{E(y_{\max})} \tag{92}$$

Since comb filter peaks are equally spaced along the frequency domain (Figure 8-2), $E(m,D_i)$ will vary considerably depending on the spectral harmonicity of the peaks of the analysed signal. Thus, Equation (92) provides a compromise between spectral harmonicity and energy filtered, which we call "spectral fit". As an example, a FIR comb filter $H(D_i,k)$ with peaks in the magnitude response matching the harmonic peaks of a monophonic signal, will have $E_m(m,D_i)$ close to 1. In contrast, a filter with peaks that do not coincide with the bins where the energy of the signal is will have $E_m(m,D_i)$ closer to 0, which is common in the onset component of a musical signal.

Since we are interested in the deviation of $E_m(m,D_i)$ from the perfect "spectral fit", the following transformation is performed:

$$E'(m,D_i) = abs(E_m(m,D) - 1) \tag{93}$$

Thus, $E'_m(m,D_i)$ equal to 0 and 1 corresponds to the perfect and worst spectral fit respectively.

### 8.2.3 Onset detection function calculation

In order to obtain the onset detection function, the changes in the spectral harmonicity are tracked. This is performed by calculating the sum of the squared difference between $E'_m(m,D_i)$ for each delay $D_i$ in each pair of consecutive frames as follows:

$$dE(m) = \sum_{i=D_{min}}^{D_{max}} \left[E'(m,D_i) - E'(m-1,D_i)\right]^2 \tag{94}$$

In Figure 8-3, the onset detection function of a tin whistle signal (top plot) obtained by utilising this approach is depicted in the middle plot. As a comparison, the energy function of the signal is also shown in the bottom plot (Equation (15)). In the ODCF onset detection function, there is a prominent peak at the onset position; however, the energy function does not show an increase in the onset component.

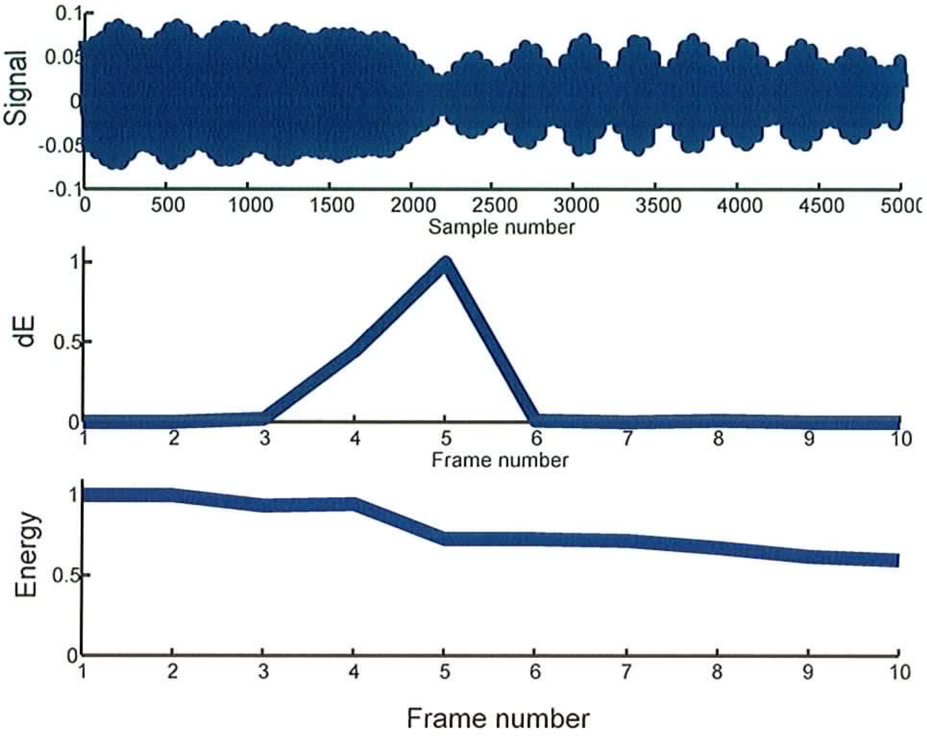

**Figure 8-3: Onset detection function of an A4-F#4 tin whistle note transition (top plot) using the ODCF method (middle plot) and Equation (15))'s energy based method (bottom plot)**

In order to illustrate how the onset peak arises, the $E'_m(m, D_i)$ function for the frame range $m$ = 2 to 6 are depicted in Figure 8-4. The delays utilised correspond to the pitch period of the 12 notes of the third octave, and the sampling frequency is 44100 Hz. It can be seen that there is not a noticeable change in the functions between frames 2 and 3, and frames 5 and 6. However, there is a significant change between frames 3 and 4, and frames 4 and 5. As can be seen in Figure 8-3, those are the frames at which the onset occurs.

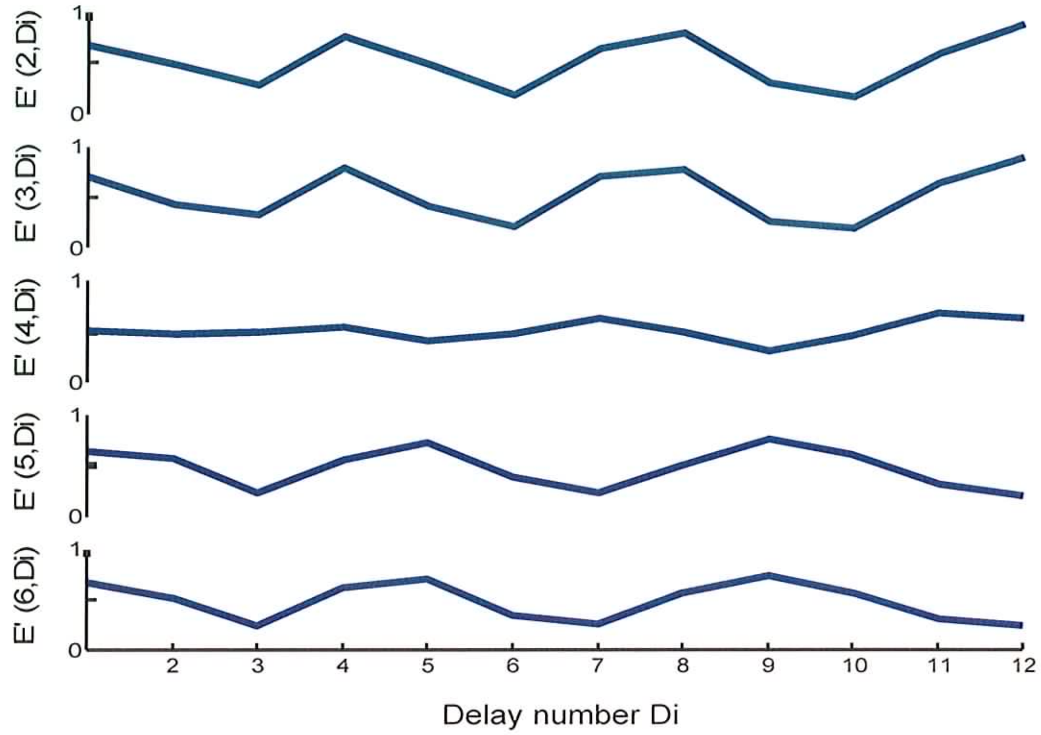

**Figure 8-4:** ***E'*(*m*,*Di*) function for frames *m* = [2...6]**

### 8.2.4 Post-processing

Other parts of an audio signal such as a slow offset – onset transition, or a part where no note is present (such as purely noise section), are also prone to spectral fit changes. In order to avoid spurious onset detections, these signal parts are detected as follows:

- ***Slow offset- onset transition***

The offset part of a signal also contains unexpected harmonicity changes, which can cause spurious onset detections. The same problem arises in the case of using phased based approaches, which are also prone to detect offsets. By contrast, energy based approaches do not produce peaks at the offset part of the signal, since this does not contain an energy increase. A possible solution would be to evaluate the onset detection function only in the sections of the signal where there is an energy increase. However, rapid note changes without an implicit energy increase will also remain undetected (Figure 8-3, bottom plot).

The proposed solution combines the ODCF with an energy based approach, which investigates the existence of energy increase peaks successively followed by an energy decrease peak. This scenario corresponds to a slow offset-onset transition. Thus, if two peaks arise in the comb filter onset detection within the above mentioned transition, the first peak is assumed to be caused by an offset event, and is discarded as an onset candidate. As an example, the top plot of Figure 8-5 depicts a tin whistle signal, where a slow offset–onset transition can be seen approximately in between samples 6000 to 10000. The comb filter onset detection function of the tin whistle signal is depicted in the middle plot, where two peaks arise during the above mentioned offset - onset transition. Finally, the bottom plot depicts a standard energy detection function, where it can be seen that a negative and a positive increase arise in the offset and onset part respectively. In this case, the first peak in the offset – onset transition is discarded as an onset candidate. Another peak also arises at around sample 1800. In this case, the onset of the new note and the offset of the previous note occur simultaneously, and consequently produce a unique peak.

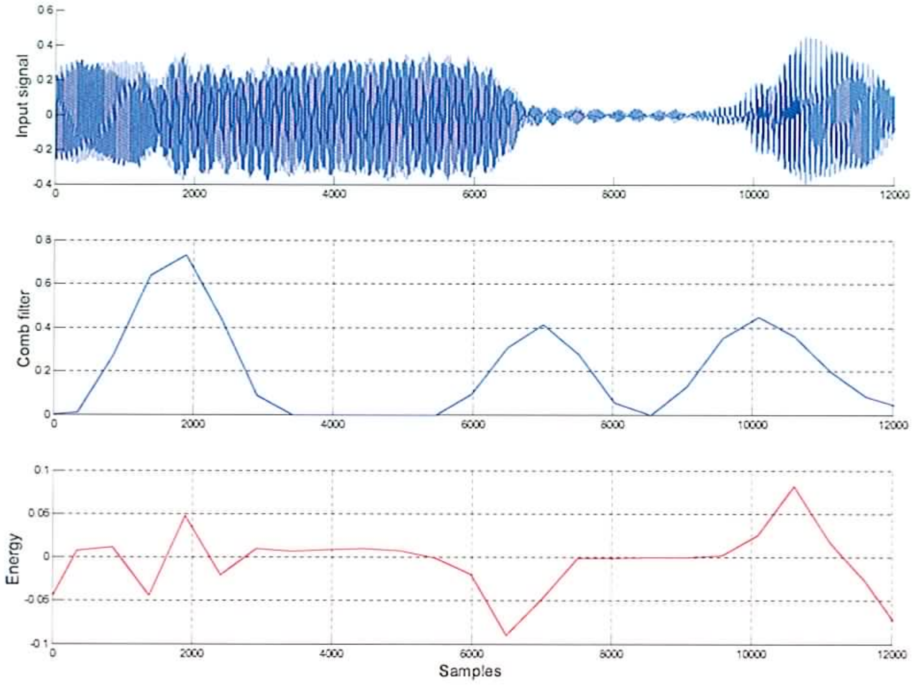

**Figure 8-5: Offset – onset transition of a tin whistle signal**

- *Noise detector*

In Equation (92), the "spectral fit" is calculated according to the energy of the analysing frame, and no distinction is made between frames containing a high or low amount of energy. Even though this technique is accurate for detecting the harmonicity in slow onset signals containing low amplitude values, the method can produce ambiguous results when analysing pure noise. In this case, the noisy frame is compounded of unpredictable values, which produce unpredictable "spectral fits", consequently producing spurious peaks in the onset detection function. In order to detect the noisy parts of the analysing signal, the following method used in [Amatriain '02] is applied:

1. The energy *E1* of the frequency range [1:3000] Hz is first obtained, which has a high concentration of energy if a harmonic signal is present in the spectrum.

2. The energy *E2* of a very high frequency range [15000:21000] Hz is also obtained. Even in the case of having a harmonic signal in the spectrum, this interval has low energy content.
3. If a note has been played, it is expected that *E1* has a much higher value than *E2*. Otherwise, the energy will be spread over the frequency axis, and it will be assumed that the signal only contains noise. Thus, by using a high threshold *Tn*, the non-noisy frames will be estimated as follows:

$$\frac{E1}{E2} > Tn \tag{95}$$

In Figure 8-6, the first 40000 samples of a tin whistle signal are depicted in the top plot. The bottom plot depicts the comb filter onset detection function of the tin whistle signal, where it can be seen that two spurious onsets due to the noise present in the signal arise. By a dashed line, the Boolean result produced by applying Equation (95) with *Tn* = 15 is depicted.

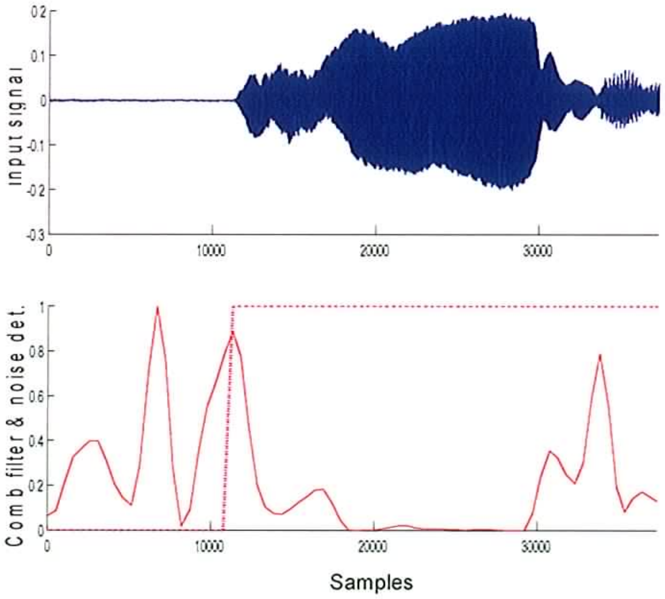

**Figure 8-6: Noise detection (bottom plot) of a tin whistle signal (top plot)**

## 8.3 Results

In order to analyse the performance of the ODCF, the onset detector has been evaluated in three different musical contexts. The first evaluation analyses the performance of the ODCF for the same tin whistle signals database used in Section 6.3. The second evaluation investigates the performance of the ODCF and the existing methods with a database of Irish traditional music instruments. Finally, the impact of amplitude and frequency modulations in the performance of the ODCF and the existing methods are also investigated.

The existing onset detection methods utilised in this section are the same as in Section 6.4: the spectral difference method [Duxbury '02], the *d*(HFC) based on [Masri '96b], the phase based method [Bello '03], and the complex based method [Duxbury '03b]. In Section 8.4, a discussion of the results obtained in this section is provided.

### 8.3.1 Evaluation of the ODCF for tin whistle tunes

As in the ODTW and the existing onset detection methods, the ODCF also allows different system parameters and configurations, which are also given by: *TFT*, *TM*, *fc* and *win* (see Section 6.3.1 for a description of the parameters). The only parameter that varies between the ODCF and the existing methods is *TFT*, which also includes the different filter delay vectors *D*.

The same testing methodology as in Section 6.3 is utilised to evaluate the ODCF performance: the first test, TEST 1, evaluates the ODCF by using different *TFT* and *TM*, where *fc* and *win* are assigned a constant value. The second test, TEST 2, configures the system with the configurations and system parameters that provide the best results in TEST 1. Next, the entire signal database is evaluated for a range of different *fc* and *win*.

The comparison of the results obtained by the ODCF in this section against the results obtained in Section 6.3 by the existing onset detection methods and the ODTW is discussed in Section 8.4.

### TEST 1: Evaluation of *TFT* and *TM*

The ODCF is evaluated for the entire database by utilising the system parameters described by *TFT* and *TM*, which are given by:

- The following filter delay vectors $Dj = [Dj_{min} \ldots Dj_{max}]$:
    - ***D3*** $= [D3_1 \ldots D3_{12}]$, which is a 12 delay vector covering the pitch period of the 12 semitones of octave 3. Thus, a filter delay $\mathrm{D3}_i = \mathrm{D3}_1$ and $\mathrm{D3}_i = \mathrm{D3}_{12}$ corresponds to the pitch period of C3 and B3 respectively.
    - ***D4*** $= [D4_1 \ldots D4_{12}]$, which is a 12 delay vector covering the pitch period of the 12 semitones of octave 4. Thus, a filter delay $\mathrm{D4}_i = \mathrm{D4}_1$ and $\mathrm{D4}_{12}$ corresponds to the pitch period of C4 and B4 respectively.
    - ***D5*** $= [D5_1 \ldots D5_{12}]$ which is a 12 delay vector covering the pitch period of the 12 semitones of octave 5. Thus, a filter delay $\mathrm{D5}_i = \mathrm{D5}_1$ and $\mathrm{D5}_{12}$ corresponds to the pitch period of C5 and B5 respectively.
- As in Section 6.3.1, the pair of system parameters *L*/*H* equal to 512/1024, 1024/2048, and 512/2048 respectively are considered.
- *TM*: As for the existing approaches in Section 6.3.2, the thresholding methods *Thres1*, *Thres2* and *Thres5* are utilised.

  As in [Beardah '95] and Section 6.3.2, the pdf in *Thres1* is generated by smoothing the histogram with a triangular kernel.

The required $\delta$ (Equation (39)) to configure *Thres5* is obtained as in Section 6.3.2 for the existing approaches. As an example, Figure 8-7 displays the percentage of *pFP* vs. *pGP* for a filter delay vector $D = D4$, and a *H/L* pair equal to 512/1024, with the best $\delta$ obtained at $\delta_{CF} = 0.04$. As a comparison, the *pFP* vs. *pGP* results obtained by using the existing approaches are also depicted (It should be noted that the *x* and *y* axis in Figure 8-7 have different scaling).

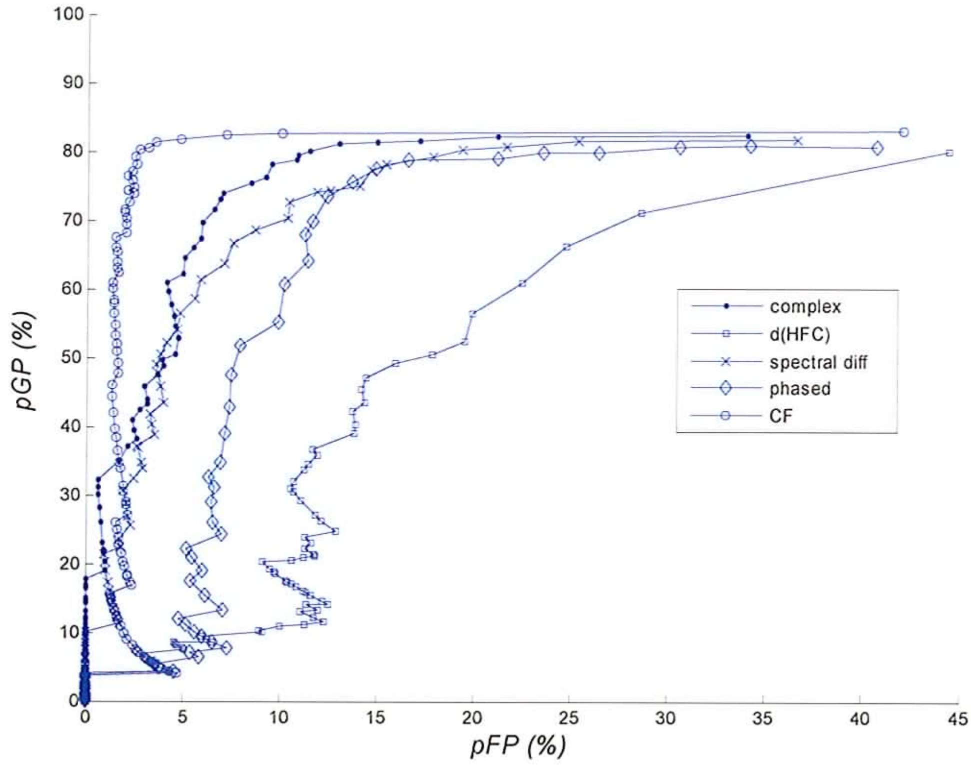

**Figure 8-7: Calculation of the best $\delta$ value for *Thres5* in the tin whistle signal database using the methods indicated in the figure legend**

The results are shown in Table 8-1, where as in Section 6.3, *win* and *fc* are set to 50 ms and (*fs*/2)*0.4 Hz respectively for all calculations.

| | | *Thres1* | | | *Thres2* | | | *Thres5* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | *D3* | *D4* | *D5* | *D3* | *D4* | *D5* | *D3* | *D4* | *D5* |
| ***L/H*** | **512/1024** | -5.68 | -8.52 | -12.17 | 52.13 | 27.79 | 35.50 | 77.28 | 77.69 | 76.06 |
| | **512/2048** | -6.90 | -11.36 | -11.36 | 61.26 | 45.44 | 37.12 | 78.09 | **80.73** | 75.46 |
| | **1024/2048** | 63.89 | 61.26 | 61.05 | 8.92 | 6.49 | 12.58 | 75.46 | 75.05 | 74.24 |

**Table 8-1: *acc* results using different delay vectors in the ODCF**

A more detailed description of Table 8-1's best *acc* result including *pGP*, *pFP* and *pFN* is shown in Table 8-2. This value is obtained for .a *L/H* pair equal to 512/2048, a filter delay *D4* and a *TM* = *Thres5*.

| | *acc* | *L* | *H* | *TM* | *pGP* | *pFP* | *pFN* |
|---|---|---|---|---|---|---|---|
| ODCF (*D* = *D4)* | 80.73 | 2048 | 512 | *Thres5* | 83.37 | 3.03 | 19.11 |

**Table 8-2: Best *acc* results and system parameters obtained by evaluating TEST 1 for the ODCF**

### TEST 2: Evaluation of *fc* and *win*

In order to obtain the best pair of *fc* and *win*, the ODCF is first configured with the set of parameters that provide the best results in TEST 1, which are shown in the second, third and fourth column of Table 8-2. Next, the best *acc* result obtained by successively using a different *win* value within its range is shown in the first column of Table 8-3. For each *win* value, the ODCF is evaluated for the entire range of *fc*.

| *win* | 20 ms | 25 ms | 30 ms | 35 ms | 40 ms | 45 ms | 50 ms |
|---|---|---|---|---|---|---|---|
| *acc* | 62.27 | 78.09 | 80.93 | 82.15 | 81.95 | 82.15 | 80.73 |

**Table 8-3: Best *acc* results obtained by evaluating the ODCF varying *win* and *fc***

In addition, a comparison of the performance of the ODCF against the existing onset detection methods by varying *fc* for a *win* equal to 50 ms is depicted in Figure 8-8.

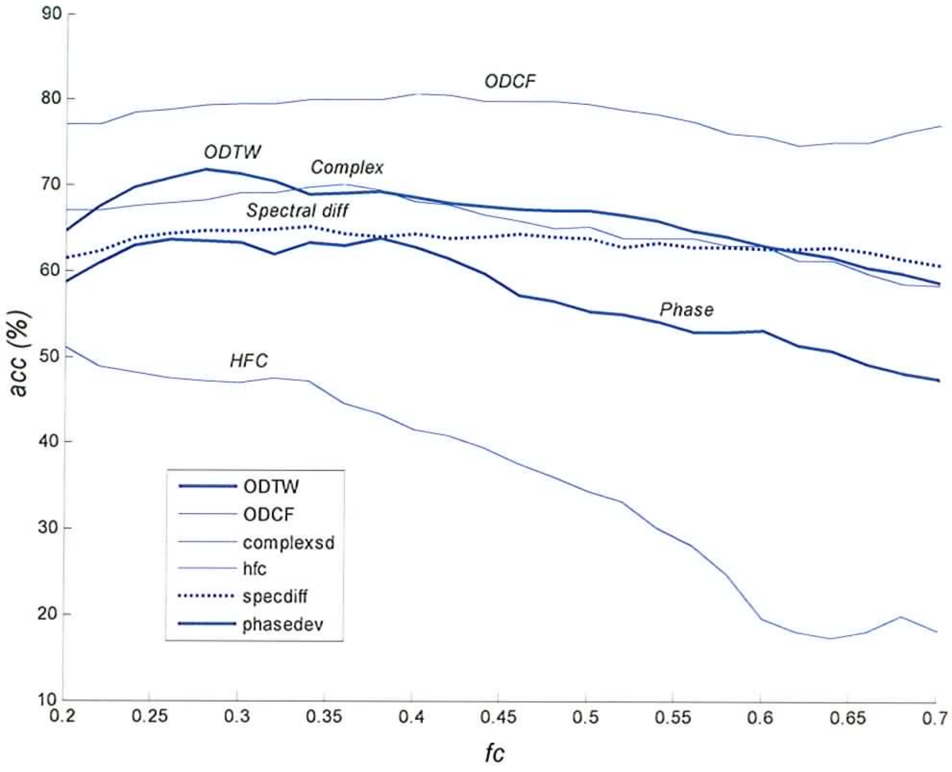

**Figure 8-8: *acc* results (*y* axis) obtained by evaluating the onset detection methods for different *fc* values (*x* axis) and a *win* = 0.05s**

### 8.3.2 Evaluation of ODCF for other slow onset instruments

As opposed to the ODTW, the ODCF can also be utilised as a general approach to detect slow onsets without requiring any prior-knowledge. In order to evaluate the performance of the comb filter approach on detecting slow onsets, a signal database of 730 hand labelled slow onsets produced by 19 excerpts of slow onset instruments is utilised. A wide range of slow traditional Irish instruments is covered by the database, which includes the flute, the accordion, the fiddle, the uilleann pipe, the susato, and slow polyphonic mixtures. As in the previous evaluations, the same tests TEST 1 and TEST 2 are performed, which will be conjointly accomplished for the existing approaches and the ODCF.

### TEST 1: Evaluation of *TFT* and *TM*

In this evaluation test, the same system parameters as in Section 8.3.1 are utilised: *D* (only for the *ODCF*), *L/H, TM, win* and *fc*. In order to configure *Thres5*, a new value of $\delta$ is obtained for each method. As an example, Figure 8-9 displays the percentage of *pGP* vs *pFP* for the ODCF and the existing approaches, where the *H/L* pair is equal to 512/1024.

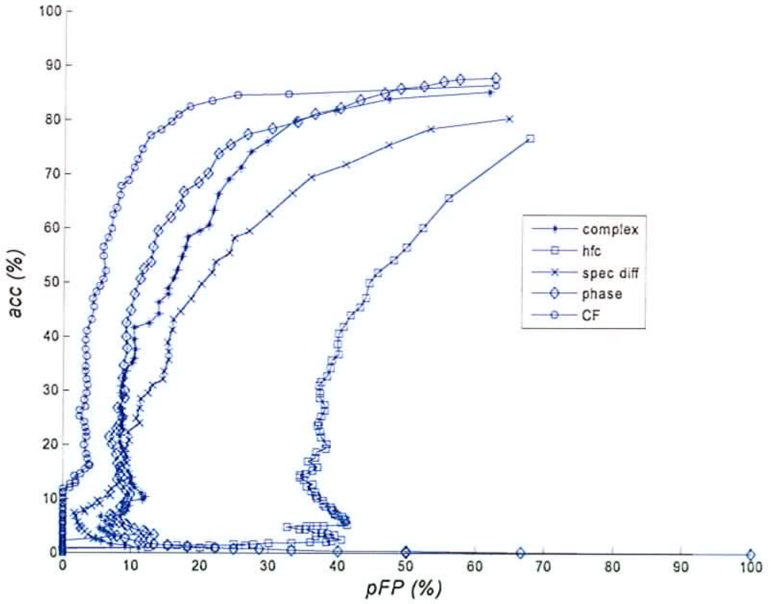

**Figure 8-9: Calculation of the best $\delta$ value for *Thres5* in the slow onset instrument signals database by using the methods indicated in the figure legend**

The best results obtained by using each onset detection method are shown in Table 8-4.

| | *acc* | *L* | *H* | *threshold* | *pGP* | *pFP* | *pFN* |
|---|---|---|---|---|---|---|---|
| ***Comb Filter* (*D* = *D4*)** | 66.80 | 2048 | 512 | *Thres5* | 81.19 | 14.86 | 19.41 |
| ***Complex*** | 45.78 | 1024 | 512 | *Thres5* | 74.06 | 27.18 | 24.94 |
| ***d(HFC)*** | 6.87 | 1024 | 512 | *Thres5* | 28.92 | 43.15 | 139.09 |
| ***Spec. diff.*** | 35.93 | 2048 | 1024 | *Thres5* | 62.91 | 29.89 | 41.09 |
| ***Phase*** | 51.23 | 1024 | 512 | *Thres5* | 75.49 | 24.16 | 24.42 |

**Table 8-4: Best *acc* results obtained by different methods in the slow onset instrument signals database.**

## TEST 2: Evaluation of *fc* and *win*

As in Section 8.3.1, the ODCF and the existing methods are first configured with the set of parameters that provide the best results in TEST 1, which are shown in the second, third and fourth columns of Figure 8-4. Then, the system is evaluated for different *fc* and *win* values within the same range as in Section 6.3. The best *acc* result for the entire range of *win* values is shown in Table 8-5.

| *win(ms)* | *ODCF* | *Complex* | *HFC* | *Spectral diff.* | *Phase* |
|---|---|---|---|---|---|
| **20** | 24.51 | -23.99 | -24.64 | -14.53 | 0.39 |
| **25** | 40.08 | -7.78 | -18.68 | -6.61 | 15.82 |
| **30** | 51.49 | 14.01 | -11.80 | 5.06 | 32.68 |
| **35** | 58.75 | 28.66 | -6.74 | 15.18 | 44.75 |
| **40** | 62.39 | 36.71 | -1.56 | 23.22 | 49.42 |
| **45** | 66.67 | 45.27 | 3.89 | 31.78 | 52.79 |
| **50** | 68.48 | 48.90 | 8.17 | 36.32 | 54.22 |

**Table 8-5: *acc* results for different *win* values in the slow onset instrument signals database**

A comparison of the onset detection methods by varying *fc* for a *win* equal to 50 ms is depicted in Figure 8-10.

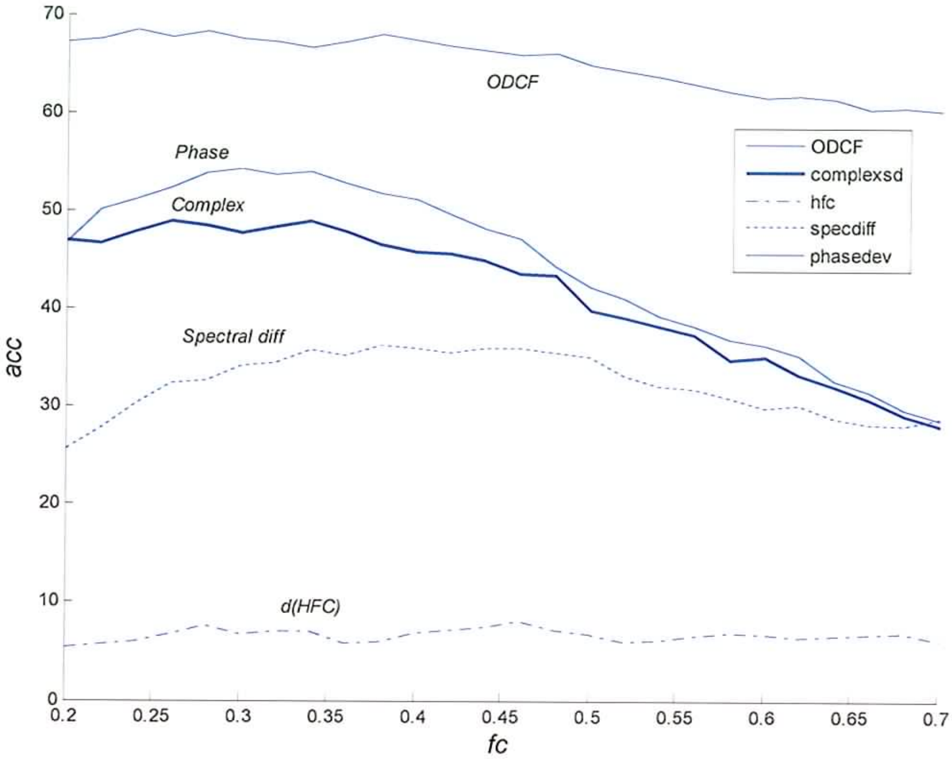

**Figure 8-10:** ***acc*** **results for different** ***fc*** **and a** ***win*** **= 50 ms in the slow onset instrument signals database**

### 8.3.3 Evaluations of the methods for Signal modulations

In Sections 8.3.1 and 8.3.2, the performance of the ODCF is evaluated for a database of tin whistle and slow onset instrument signals. As described in Chapter 4, frequency and amplitude modulations affect the performance of existing onset detectors. In order to recreate an amplitude modulation scenario, a signal is synthesised by utilising the audio editing tool "Cool Edit Pro" [Cool '02]. A C3 tone with 5 harmonics is first synthesised.

Then, the signal is modulated by an envelope. The resulting signal is depicted in Figure 8-11, plot A. The onset detection function generated by the ODCF is depicted in plot F, and the resulting functions by using the existing onset detection methods are depicted in plots B to E.

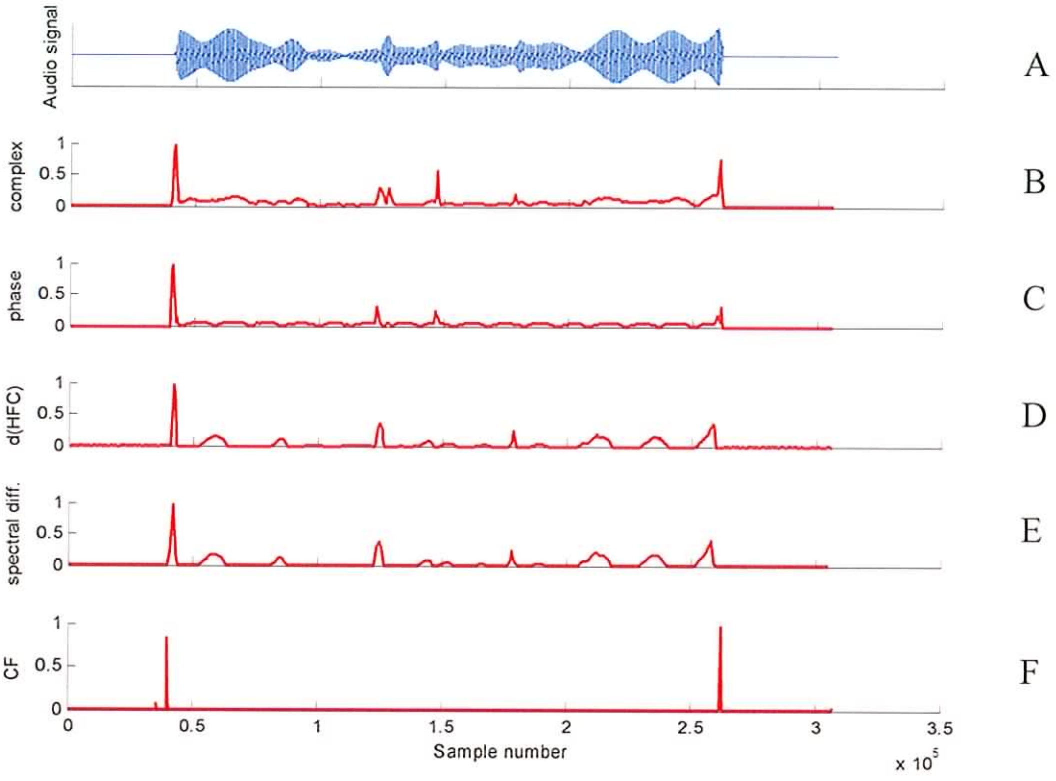

**Figure 8-11: Onset detection functions of the amplitude modulated synthetic signal depicted in plot A by using different techniques, which are labelled in the *y* axis of plots B to F**

Apart from amplitude modulations, frequency modulations can also arise in the signal, which consequently affect the onset detection accuracy. In Figure 8-12, the onset detection function of the same tin whistle signal as in Figure 6-8 is depicted in the bottom plot. The middle and top plots depict the waveform and the spectrogram of the signal respectively, where the amplitude and frequency modulations that arise in the signal can be seen. The E5 note depicted in Figure 8-12 is played using a slide effect, which inflects the pitch to reach F5#, which means that a modulation between approximately 659 Hz to 740 Hz occurs. The onset detection functions obtained by using the existing onset detection methods have previously been depicted in Figure 6-8.

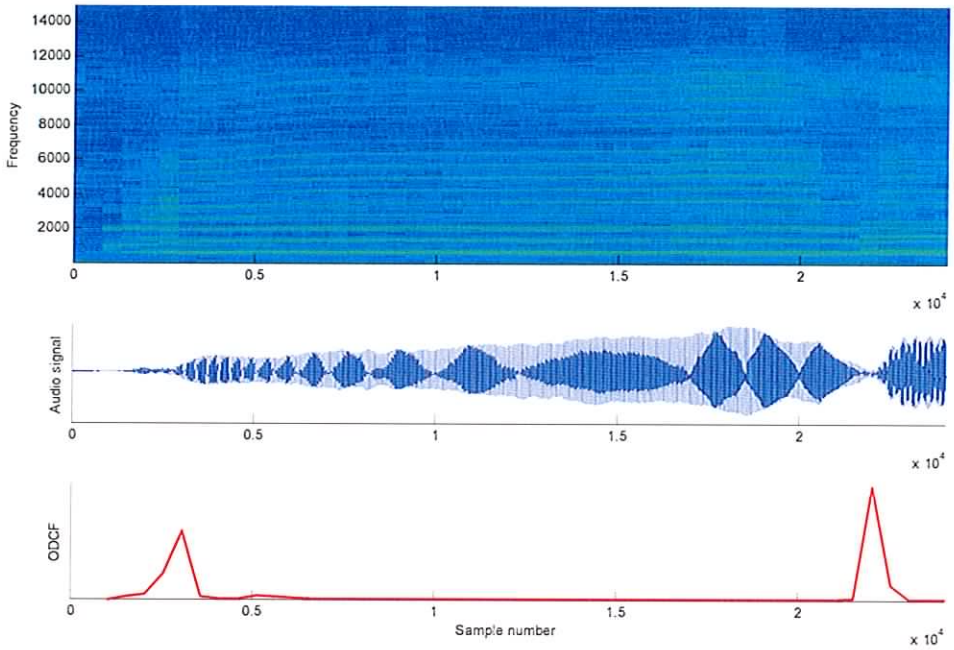

**Figure 8-12: Onset detection function by using the ODCF (bottom plot) of a tin whistle signal playing E5 using a slide effect (middle plot), whose spectrogram is depicted in the top plot**

## 8.4 Discussion

In Section 8.3, the performance of the ODCF has been evaluated for a range of system parameters and configurations, which includes the comb filter delay vector *D*, the thresholding methods *TM*, the pair *H*/*L* (hop/frame), the LPF cut-off frequency value *fc* and the window length *win*. The results are compared against the performance of existing approaches, which are also evaluated for the same parameters except the filter delay vector *D*. The methods have been compared by using two different databases of audio signals, and also with amplitude and frequency modulated signals. The first database comprises tin whistle signals, and the second one slow onset instrument signals used in Irish traditional music.

As it can be seen in Table 8-1 and Table 8-4, by evaluating the ODCF for both databases, *Thres5* is the best performing threshold method. This result is not surprising considering that *Thres5* utilises the knowledge of the location of the onsets to provide the best threshold value. By using *Thres5*, the vector delay *D4* provides the best results for both databases. However, it can be seen that by using *D3* or *D5* the results do not vary significantly. This is explained by analysing the comb filter spectral shape, whose magnitude response has a harmonic structure. A comb filter built for a given semitone of octave *i*, is also harmonically related to higher octaves; 2:1 with octave *i*+1, 4:1 with octave *i*+2 and so on (Figure 5-6). Thus, the harmonicity of higher octave notes is also tracked by lower octave comb filters.

Table 8-1 also shows that by using *Thres5* there is not any pair of *L*/*H* that provides substantially better results than the other pairs. However, a long hop size such as *H* =1024 samples will be prone to spurious detections, since the signal spectral fit can change during that interval as is the case in a frequency modulation. Since the spectral fit is calculated based on the energy content of the frame, small frame length values, such as *L* = 1024, can also be oversensitive to signal changes. Consequently, the pair *L/H* equal to 512/2048 provides a good compromise between robustness and sensitivity to signal changes.

The set of parameters that produce the best *acc* results by evaluating TEST 1 (which maintains a constant value *fc* and *win*) using the ODCF for the tin whistle signals database are shown in Table 8-2. By comparing these results against the results that the ODTW and the existing methods produce during the same test (Table 6-4), it can be seen that the two best performing methods are the ODCF and the ODTW, whose best *acc* values are equal to 80.73 and 68.56 respectively. This result reflects the improvement of the ODCF upon existing onset detection methods.

Table 8-4 shows the results obtained by evaluating the ODCF and the existing approaches using TEST 1 for the slow onset instrument signals database. In this case, the two best performing methods are the ODCF and the phase based approach, whose best *acc* values are equal to 66.80 and 51.23 respectively. This result shows that the ODCF also improves upon existing onset detection methods for other slow onset instruments.

The result that varies more between the ODCF and the rest of the methods in both databases is *pFP*, whose value is significantly lower in the ODCF (Table 8-2, Table 8-4 and Table 6-4). This is due to various factors; firstly, the post-processing tasks (noise and offset-onset transition detection) reduce spurious onset detections. In addition, the ODCF is capable of dealing with frequency and amplitude modulations, by producing onset detection functions that remain stable during the harmonic part of the signal. This is shown in Section 8.3.3, where the performance of the ODCF and the existing methods is evaluated in the context of amplitude and frequency modulations. The performance of the

ODCF for the amplitude modulation test is depicted in plot F of Figure 8-11, where it can be seen that the onset detection function shows two distinctive peaks in the onset and offset part, and that the amplitude modulations do not affect the performance of the system. Since the ODCF method relates the harmonicity to the energy of the frame, a change in the signal energy between frames without an implicit change in the harmonicity, is not noticed in the onset detection function. Existing energy based approaches, such as the spectral difference and the *d*(HFC) methods (plots D and E respectively), encounter difficulties when dealing with strong amplitude modulations.

Figure 6-8 and Figure 8-12 depict the onset detection functions obtained by analysing an amplitude and frequency modulated tin whistle signal. By comparing both figures, it can be seen that, as opposed to the ODTW and the existing approaches, the frequency modulation does not alter the performance of the ODCF. It can be seen in the top plot of Figure 8-12 the frequency modulation occurs gradually, which does not substantially change the harmonicity detection value of the comb filters if using small hop sizes (e.g., $H = 512$ samples).

Another manner of comparing the performance of the onset detection systems that use the thresholding method *Thres5*, is by analysis of the figures that display the percentage of *pFP* vs. *pGP* for different $\delta$ values. This measure is depicted in Figure 8-7 and Figure 8-9 for the two signal databases, where the *H*/*L* pair equal to 512/1024 is utilised in the example. As opposed to the phase and complex based approaches, the *H*/*L* pair equal to 512/1024 is not the best performing pair in the ODCF (Table 8-2 and Table 8-4).

However, it can be seen that even by using this *H*/*L* pair, the ODCF line of results is the closest to the top-left corner in both figures.

As can be shown by the onset detection methods for the tin whistle signals database (Table 6-4 and Table 8-2) and with the slow onset instrument database (Table 8-4), it can be seen that the performance of all the onset detection methods degrade in the slow onset instrument database. This shows the difficulty of having an onset detection system, and a thresholding method that works robustly for all the musical contexts.

The ODCF has also been evaluated for different *fc* and *win* values (TEST 2). The performance of the ODCF by utilising different *win* values is shown in Table 8-3 and Table 8-5 for the tin whistle and the slow onset instruments signal database respectively. Table 8-3 shows that the ODCF for the case of a *win* equal to 30 ms performs better than the ODTW and the existing approaches for the most tolerant *win* equal to 50ms (Table 6-5). This is also the case in Table 8-5, where by using a *win* equal to 35 ms, the ODCF improves all the existing methods for a *win* equal to 50ms. This shows that the onset time accuracy provided by the ODCF improves upon the existing onset detection methods. In addition, as shown by the ODCF results of Table 8-3, the best *acc* results are not provided by the most tolerant *win*. It should be recalled that apart from setting the tolerance in the onset detection results comparison, the parameter *win* is also used to combine very close peaks that are assumed to belong to the same onset. However, successive ornamentation and onset events that are separated by less than *win* are also combined into one unique candidate, which can result in a missed event detection. The

ODCF is capable of dealing with this scenario, even with the drawback of using a shorter *win* in the comparison between the results obtained and the real location of the onset. Another illustrating comparison between methods is depicted in Figure 8-8 and Figure 8-10, where the methods are evaluated for the entire *fc* range and a *win* equal to 50 ms for the two signal databases respectively. In both figures, the line of results representing the ODCF provides the best results in the entire *fc* range.

## 8.5 Conclusions

A novel onset detector system based on comb filters has been presented in Section 8.2, which focuses on the harmonic characteristics and energy changes of the audio signal. Both properties are combined by utilising FIR comb filters on a frame by frame basis in order to obtain an onset detection function, which is suitable for detecting slow onsets, and for signals that have amplitude and frequency modulations. In addition, post-processing tasks such as the detection of noise and offset-onset transition sections have been developed in order to reduce the number of spurious onset detections. The development of this novel onset detection approach corresponds to Contribution 3, as outlined in Section 1.2.1.

The onset detector has been evaluated by using two different databases, which comprise tin whistle tunes and other Irish traditional music instrument tunes respectively. The results presented in Section 8.3 and discussed in Section 8.4 improve upon the existing approaches.

The ODTW also improves upon certain limitations of the ODTW presented in Chapter 6, by providing both a more accurate onset time and by dealing with strongly modulated signals in amplitude. Gradual frequency modulations that arise in the signal, as in the case of a slide effect are adequately catered for by the ODCF. However, the system is vulnerable to rapid frequency modulations. The investigation of a method for overcoming these limitation warrants future research.

In the following chapter, another system based on comb filters is presented. The approach aims to detect the harmonic accompaniment provided to melodic instruments in Irish Traditional music.

# 9 Multi-pitch Estimation Using Comb Filters (MPECF)

## 9.1 Introduction

In Chapter 8, an onset detector based on comb filters (ODCF) was presented. The results provided by the system show that the harmonic type of magnitude response of comb filters is a very useful feature for dealing with Irish traditional music harmonic instruments. In recent years, harmonic accompaniment has been added to Irish traditional music. As discussed in Chapter 5, periodicity based pitch detection methods are less efficient when there is more than one source present in the signal, as it occurs when musical accompaniment is utilised. In this case, sharper transition band magnitude responses are required to perform the filtering.

As in the ODCF (Chapter 8), the MPECF based the system on comb filter techniques. The approach is based on the [Tadokoro '03] method, which generates a notch comb filter per note of the considered pitch detection range. When this method detects a new note, its harmonics are automatically extracted from the spectrum during the same operation, automatically generating a residual signal. Following this, the detection of the remaining existing notes can be performed. In addition, this method provides the flexibility to connect a mode detector to the pitch detection system, which will reduce the number of comb filters required in the detection. In the MPECF, the FIR comb filters utilised by [Tadokoro '03] are replaced by another type of comb filter, which alters the remaining spectrum after filtering to a lesser degree. In addition, the magnitude response of this filter is weighted, which reduces the number of octave pitch detection errors. A method to detect the harmonic triads provided by the accompaniment and the notes played by the leading instrument is also provided. The development of the system represents Contribution 4 in Section 1.2.1.

In order to accompany Irish Traditional music, the tune and its mode should be first known [James '02, McQuaid '05]. In [James '02], a document produced by a collective collaboration of Irish traditional players on the subject of accompanying Irish traditional music is shown. They produce lists of chords organised by modes, where each list is sorted by most to least chords played [James '02]. The first six chord names of each list provided by the players only comprise of minor or major triads. In addition, [James '02] states that the first four chords of each list are generally enough to provide the harmonic accompaniment. Consequently, the problem of transcribing harmonic accompaniment is reduced into the detection of major and minor triads. This shows that harmonic accompaniment in Irish traditional music is relatively simplistic [Carolan '06].

In Section 9.2, the multi-pitch estimator is first introduced. A set of results that evaluate the pitch detection system are presented in Section 9.3. Next, a discussion of the results is provided in Section 9.4. Finally, conclusions regarding the multi-pitch estimation system are provided in Section 9.5.

## 9.2 MPECF system description

In order to transcribe the musical triads played by the harmonic accompaniment, a system based on [Tadokoro '03] model is utilised (Section 5.2.1), which is depicted in Figure 9-1. As in [Tadokoro '03], the MPECF filter that produces an amplitude minimum represents the first detected note. Next, other notes in the audio signal are detected by iteratively connecting the output of the filter that has produced the minimum with the input of the parallel comb filter system [Tadokoro '03]. The same filtering process is repeated again until all the notes have been extracted. After estimating the notes, an existing major or minor triad present is transcribed.

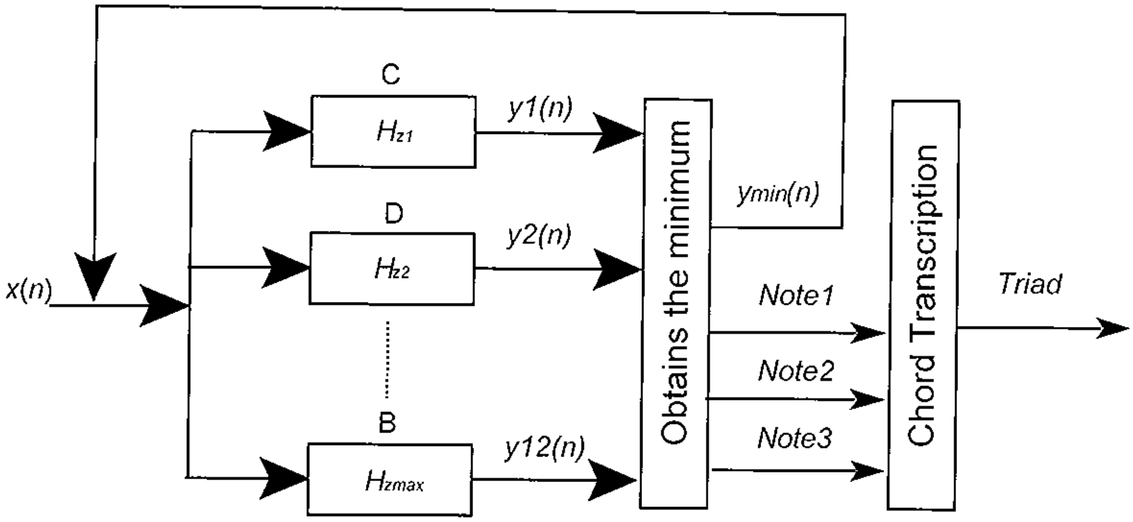

**Figure 9-1: MPECF system for triad detection**

In the MPECF, the bank of FIR comb filters utilised by [Tadokoro '03] are replaced by a bank of a different type of comb filter, whose characteristics are described in Section 9.2.1. Next, Section 9.2.2 describes the need to weight the magnitude response in order to avoid low octave pitch detection errors. Finally, Section 9.2.3 provides the method to transcribe triads based on the minima produced by the comb filters.

### 9.2.1 Comb Filter characteristics

As shown in Chapter 5, the performance of periodicity based methods such as comb filters or the autocorrelation method degrade when analysing polyphonic signals. For example, FIR comb filters have a wide bandwidth around the notch frequencies, which attenuates the remaining spectrum after filtering. In order to avoid the signal amplitude alteration caused by FIR comb filters, a zero for every pole can be added into the frequency response [Proakis '95]. This method has been utilised for eliminating harmonic interference signals in electrocardiograms (EEG) measures [Chang-Tar '94; Soo-Chang '97], or to eliminate the power line of 60 Hz and its harmonics in instrumentation [Proakis '95]. Thus, the frequency response of the modified comb filter becomes:

$$H(z) = \frac{1 - Z^{-D}}{1 - \rho D^{-D}} \qquad (96)$$

where $0 < \rho < 1$ is a stability condition.

The closer $\rho$ is to 1, the closer the poles are to the unit circle, and the narrower the notches will be. The frequency response of the comb filter is depicted Figure 9-2, where it can be seen that the filter pass-band is very flat.

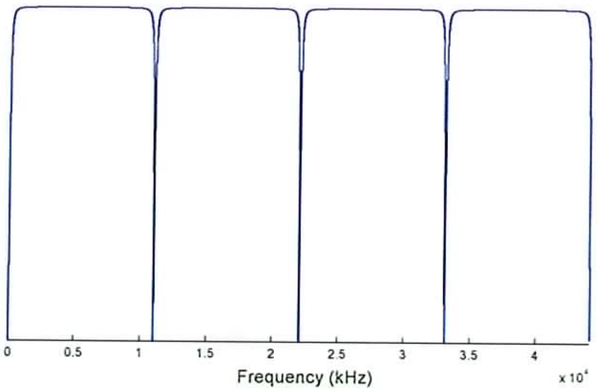

**Figure 9-2: Equation (96) comb filter magnitude response**

In order to preserve the remaining spectrum after filtering the harmonics of the detected note, Equation (96) comb filter is incorporated into a polyphonic transcription context [Gainza '05a]. This filter can be interpreted as a combination between a standard FIR Comb filter and a standard IIR Comb filter with transfer functions $H_l$(z) and $H_2$(z) respectively:

$$H(z) = 1 - Z^{-D} \tag{97}$$

$$H_2(z) = \frac{1}{1 - \rho D^{-D}} \tag{98}$$

The poles in $H_2$(z) produce resonances around the frequency of the notch, flattening the pass band and sharpening the filter stop band. This is illustrated in Figure 9-3, where $H_l$(z) and $H_2$(z) are depicted by the solid line and dashed line respectively, and where $D$=100 samples and $\rho$ = 0.6.

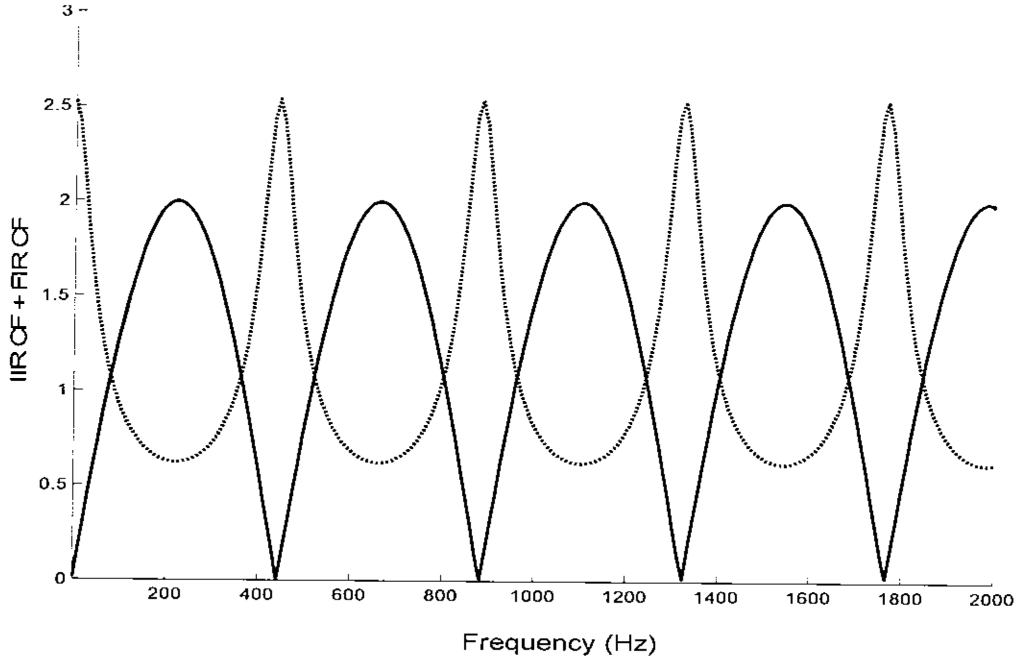

**Figure 9-3: Equation (96) comb filter (dashed line) and FIR comb filter (solid line)**

As an example, Figure 9-4 depicts the magnitude response of Equation (96) comb filter with $\rho = 0.6$ and $D = 100$ (dashed line) and the same polyphonic signal of Figure 5-10. A notch FIR comb filter is also depicted by a solid line. It can be seen that the IIR comb filter nulls coincide with the frequencies of the harmonic peaks of A4, and alter the amplitude of the C5 harmonics to a lesser degree than the FIR comb filters.

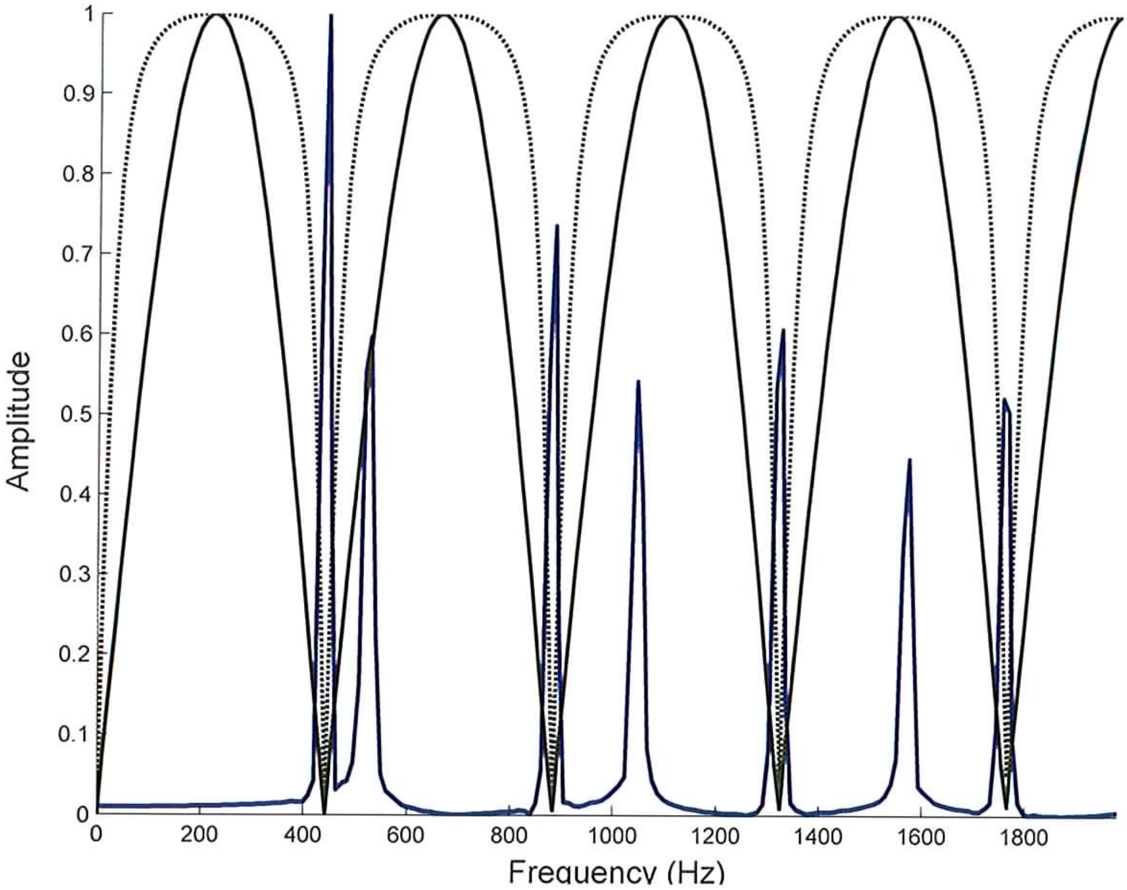

**Figure 9-4: Equation (96) comb filter (dashed line) and FIR comb filter (solid line) filtering of a polyphonic signal,**

### 9.2.2 Weighting the comb filter magnitude response

- **Problem of detecting chords using comb filters**

Consider that the fundamental frequency of a note N1 and a note N2, denoted as $f_1$ and $f_2$ respectively, are related thus:

$$f_1 = \frac{m}{n} \times f_2 \qquad (99)$$

where $m$ and $n$ are integers.

Every $n^{th}$ harmonic of $N_1$ overlaps a corresponding $m^{th}$ harmonic of the sound $N_2$.

The degree of overlapping between the notes depends on the musical interval. As an example, D and A form a perfect fifth ($n/m$=3/2), where every third harmonic of D overlaps every second harmonic of A. If more than two notes are present in the polyphony, the number of overlapped harmonics in a note could be greater than the number of non-overlapped harmonics [Klapuri '98]. As described in the introduction of this chapter, harmonic accompaniment in Irish Traditional music is generally provided by major and minor triads [James '02]. In Table 9-1, examples of frequency relations between the notes comprising common triads are shown. The first two examples correspond to two widely utilised triads in Irish Traditional music: D major and E minor. The frequency relation between notes is shown related to the triad root, where it can be seen that in the D major triad, the root D is in a 5:4 and 3:2 relation to F# and A. Thus, half (3/6) of the harmonics of D are overlapped with other note harmonics. By analysing the frequency relations from A, this note is in 5:6 and 2:3 relation to D and F# respectively. Thus, more than half (4/6) of the harmonics of B overlap with other notes. The same principle applies to any major triad root.

| Chord | Chord note | | | | *m:n* relation to N1 | | | |
|---|---|---|---|---|---|---|---|---|
| | N1 | N2 | N3 | N4 | N1 | N2 | N3 | N4 |
| **D major triad** | D | F# | A | | 1 | 5:4 | 3:2 | |
| **E minor triad** | E | G | B | | 1 | 6:5 | 3:2 | |
| **D3 + D4 Major triad** | D3 | D4 | F4# | A4 | 1 | 2 | 5:2 | 3 |
| **D2 + D4 Major triad** | D2 | D4 | F4# | A4 | 1 | 4 | 5 | 6 |

**Table 9-1: Example of frequency relations between chord notes (adapted from [Klapuri '98])**

As described in Chapter 5, time domain periodicity based methods such as comb filter techniques are prone to low octave pitch detection errors. By considering the third chord example of Table 9-1, which comprises a D4 major triad and a D3 note (one octave lower than D4), it can be seen that all the harmonics of D4 and A4 overlap with the harmonics of D3. In addition, half of the harmonics of F#4 overlap D3 harmonics. The fourth chord example of Table 9-1 shows a D2 (two octaves lower than D4), playing with the same D4 major triad. It can be seen that all the harmonics of D4, F#4 and A4 overlap the harmonics of D2, which signifies that a comb filter built for note D2 will extract all the harmonics of D4, F#4 and A4, even without the presence of D2. This shows that avoiding low octave pitch detection errors when analysing triads by using comb filters is crucial.

As described in Chapter 5, by limiting the amount of peaks in the comb filter structure, low octave pitch detection errors are reduced. However, the system becomes more prone to high octave pitch detection errors [Brown '92]. A compromise can be obtained by weighting the comb filter magnitude response. This method is utilised in [Martin '82, Morgan '97] to detect the pitch of the speech, by generating comb filter peaks whose amplitude decrease with the frequency. This method allows utilising a large number of comb filter peaks, but giving more weight to low harmonics. In addition, by weighting the comb filter magnitude response, a given amount of energy of the overlapped harmonics remains in the residual spectrum after filtering, which affects the detection of subsequent notes to a lesser degree.

- **Applying weighting to the comb filter magnitude response**

As opposed to [Martin '82, Morgan '97] comb filters, Equation (96) comb filters based the pitch detection on producing filter output energy minima. Thus, in order to apply the weighting to the magnitude response, the following method applies:

1. The comb filters are spectrally reversed. As a result, the comb filter notches and pass bands become peaks and stop bands respectively [Smith '97].
2. A weighting function is applied to the spectrum [Martin '82, Morgan '97]. Thus, the comb filter peak values decrease with frequency.
3. The resulting weighted spectrum is again spectrally reversed. Consequently, the comb filters notch amplitude increase with frequency.

Figure 9.5 depicts an example of a weighted modified comb filter, where it can be seen that the weight given to the magnitude response decreases with the comb filter notch number. In musical terms, energy of the harmonics of a note whose period coincides with the comb filter delay is extracted, where the amount of energy extracted decreases with the harmonic number.

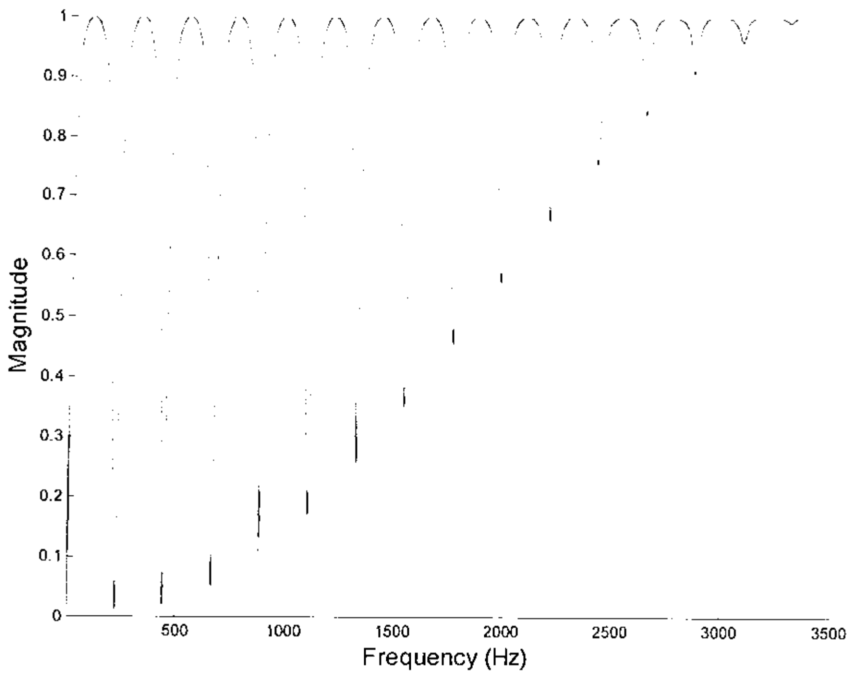

**Figure 9-5: Weighted comb filter**

### 9.2.3 Triad Transcription method

- **Transcribing the accompaniment triad**

From Figure 9-1, the MPECF iteratively produces filter output minima, estimating notes *N1*, *N2* and *N3*, which correspond to filters [*i*(*N1*), *i*(*N2*) , *i*(*N3*)] respectively.

As an example, if octaves 2, 3 and 4 are considered, 36 comb filters *Hz*(*i*) are built (where *i*=1...36). Thus, comb filter output minima at *I* = [24, 27, 32], correspond to B3, D4 and G4 respectively.

Next, the root of the triad is obtained as follows:

In order to comprise all the values of *I* = [*i*(*N1*) , *i*(*N2*) , *i*(*N3*)] within the 1:12 range (which corresponds to the C to B range), the remainder of the division *R*=[*r*(*N1*), *r*(*N2*), *r*(*N3*)] = [*i*(N1): *i*(*N2*) :*i*(*N3*)]/12 is obtained. If a component of R vector is equal to 0, a value equal to 12 is assigned. Thus, $R$ in the above example will be equal to $R$ = [24, 27, 32]/12 = [12, 3, 8].

Following this, all possible permutations of $R$ are computed, where the first position in the permuted vector $R$ represents the triad root candidate $C$. Next, components of the permuted vector with a value smaller than the root C, are added a value of 12, which is equivalent to increase the vector component one octave. For example, the permutation of $R$ equal to [8, 12, 3], where $C$=8, is converted to [8, 12, 15].

Finally, the triads are transcribed; if the permutation [$r(N1)$, $r(N2)$, $r(N3)$], where $C$= $r(N1)$, follows the pattern:

- [$r(N1)$: $r(N1) + 4$: $r(N1) + 7$], an $N1$ major triad is obtained.
- [$i(N1)$: $i(N1) + 3$: $r(N1) + 7$], an $N1$ minor triad is obtained.

Following the above example, the permutation [8, 12, 15] produces a G major triad, which comprises notes G, B and D.

- **Transcribing the accompaniment triad and the melodic instrument note**

If both accompaniment and leading melodic instrument are playing within the same octave, the task of transcribing both triad and note played by a melodic instrument becomes more difficult. In the case of both leading instrument and accompaniment playing the same note, 100% of the leading instrument harmonic notes overlap accompaniment instrument harmonics. If both instruments play a different note within the same octave, it is difficult to know the instrument that has produced each of the notes. In this case, the features of each note should be extracted from highly overlapped harmonics to enable an instrument recognition algorithm classify the detected notes [Eronen '00].

However, this is not always the case in Irish traditional music. As an example, the D key tin whistle plays in octaves 5 and 6, as opposed to instruments such as guitar and bouzouki which generally play the accompanying triads in lower octaves. In addition to playing in a different octave range, the leading instrument generally is played louder than the accompaniment instrument. Consequently, even in the case of both instruments playing the same note, the harmonics of the leading instrument remain more prominent. A method for detecting both triads and melodic notes is shown in Figure 9-6. Two different filter banks for the leading and accompaniment instruments are utilised to cover their respective octave ranges (e.g., octaves 5 and 6 for the D key tin whistle, and octaves 2, 3 and 4 for the guitar). Firstly, the output of the filter that produces the first minimum, by using the filter bank designed for the octave range of the leading instrument, provides the note played. This output is connected to the input of the triad detector of Figure 9-1, which detects the triad played by the harmonic accompaniment.

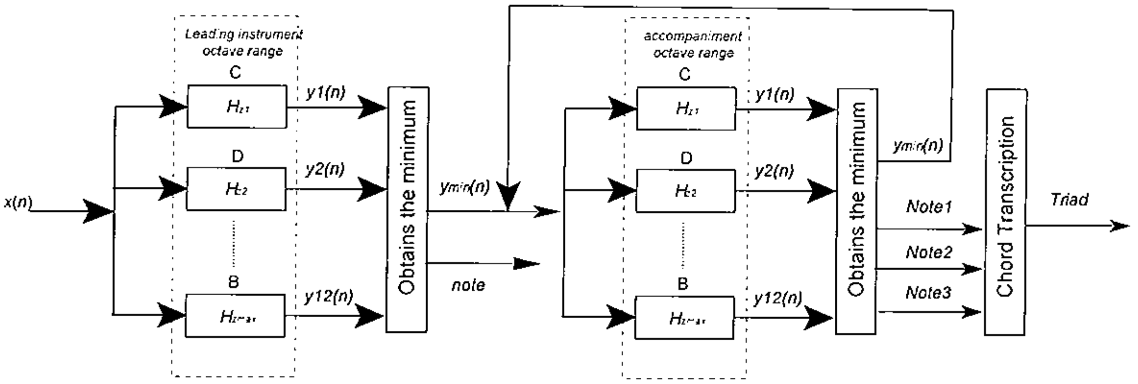

**Figure 9-6: Accompaniment and leading instrument transcription using the MPECF.**

## 9.3 Results

In order to evaluate the performance of the algorithm, two different set of tests have been performed. Firstly, Section 9.3.1 compares the MPECF against the multi-pitch transcription model of the [Tadokoro '03] model by using a set of synthetic signals. Next, the performance of the MPECF is evaluated using real guitar triads and tin whistle signals in Sections 9.3.2. The results obtained in this section are discussed in Section 9.4.

### 9.3.1 Comparison of multi-pitch systems

In this section, the result of replacing the bank of FIR comb filters utilised by the [Tadokoro '03] model by the comb filter introduced in Equation (96) is investigated. A range of monophonic and polyphonic noise-free signals are synthesised. Each synthetic note is comprised of five harmonics with amplitude ratio equal to the fundamental harmonic amplitude divided by the harmonic number. In order to avoid octave pitch detection errors in the comparison, the audio signals are restricted to octave 4, and both methods build 12 note comb filters only for the same octave 4. Equation (96) comb filter is generated using a $\rho = 0.7$ , which is obtained experimentally. The pitch detection analysis has been performed by windowing the signals using two different frame lengths $L$ equal to 4096 and 8192 samples respectively.

The comb filter methods are evaluated for monophonic signals (TEST 1), two note polyphonic signals (TEST 2), three note polyphonic signals (TEST 3) and four note polyphonic signals (TEST 4).

### TEST 1: Monophonic Signals

The test consists of a database of the 12 semitones of the fourth octave. The results for both systems are shown in the first row of Figure 9-2, where it can be seen that both systems correctly detect all the notes for both frame lengths.

| | MPECF | | [Tadokoro '03] | |
|---|---|---|---|---|
| | *L*= 8192 | *L*= 4096 | *L*= 8192 | *L*= 4096 |
| TEST 1 | 12/12 = 100 % | 12/12 = 100 % | 12/12 = 100 % | 12/12 = 100 % |
| TEST 2 | 11/11 = 100 % | 11/11 = 100 % | 10/11 = 90.9 % | 7/11 = 63.63 % |

**Table 9-2: Monophonic signals (TEST 1) and two note polyphonic signal (TEST 2) results**

### TEST 2: Two Note Polyphonic mixture

All the realisable two note mixtures derived from the note C are analysed [CC, CD, CD#, CE, CF, CF#, CG, CG#, CA, CA#, CB]. The results can be seen in the second row of Table 9-2. The MPECF detects all the notes, and [Tadokoro '03] model provides 90.9% and 66.63% of correct results using a frame length *L* equal to 8192 and 4096 samples respectively.

### TEST 3: Three Note Chords (Triads)

The detection of four common three note chords is investigated in this test. The MPECF again correctly detects all notes present in the polyphony and [Tadokoro '03] model provides 83.3 % and 58.33 % of correct results using a frame length *L* equal to 8192 and 4096 samples respectively.

| **Test signals** | **MPECF** | | **[Tadokoro '03]** | |
|---|---|---|---|---|
| | ***L*= 8192** | **L= 4096** | ***L*= 8192** | **L= 4096** |
| C,E,G (C Major) | 3/3 = 100 % | 3/3 = 100 % | 2/3 = 66.6 % | 2/3 = 66.6 % |
| C,D#,G (C Minor) | 3/3 = 100 % | 3/3 = 100 % | 2/3 = 66.6 % | 1/3 = 33.3 % |
| C,E,G# (C Augmented) | 3/3 = 100 % | 3/3 = 100 % | 3/3 = 100 % | 2/3 = 66.6 % |
| C,D#,F# (C Diminished) | 3/3 = 100 % | 3/3 = 100 % | 3/3 = 100 % | 2/3 = 66.6 % |
| Total | 12/12 = 100 % | 12/12 = 100 % | 10/12 = 83.3 % | 7/12 = 58.33 % |

**Table 9-3: Three note chord detection results (TEST 3)**

## TEST 4: Four Note Chords

In the context of four note chords, the MPECF detects correctly all the chord notes of Table 9-4. However, [Tadokoro '03] model provides 85 % and 30 % of correct results using a frame length *L* equal to 8192 and 4096 samples respectively.

| **Test signals** | **MPECF** | | **[Tadokoro '03]** | |
|---|---|---|---|---|
| | ***L*= 8192** | ***L*= 4196** | ***L*= 8192** | ***L*= 4196** |
| C,E,G,A (C Major 6) | 4/4 = 100 % | 4/4 = 100 % | 3/4 = 75 % | 0/4 = 0 % |
| C,E,G,A# (C Dominant 7) | 4/4 = 100 % | 4/4 = 100 % | 4/4 = 100 % | 2/4 = 50 % |
| C,E,G,B (C Major 7) | 4/4 = 100 % | 4/4 = 100 % | 3/4 = 75 % | 2/4 = 50 % |
| C,D#,G,A (C Minor 6) | 4/4 = 100 % | 4/4 = 100 % | 3/4 = 75 % | 1/4 = 25 % |
| C,D#,G,A # (C Minor 7) | 4/4 = 100 % | 4/4 = 100 % | 4/4 = 100 % | 1/4 = 25 % |
| Total | 20/20 = 100 % | 20/20 = 100 % | 17/20 = 85 % | 6/20 = 30% |

**Table 9-4: Four note chord detection results (TEST 4)**

### 9.3.2 Detection of guitar triads

The MPECF is evaluated for real signals by performing two different tests. Firstly, TEST 1 evaluates the system for 13 major and minor guitar triad signals. Then, the system is evaluated for 14 signals containing mixtures of guitar triads accompanying tin whistle notes. In order to investigate the impact of the location of the analysing frame in the accuracy of the MPECF, a sliding Hanning window with length $L$ =4096 is applied to the audio signal. The frame varies its position in the signal by using a hop size equal to 2048 samples.

**TEST 1: Detection of Guitar triads**

The database of signals utilised comprise 13 major and minor guitar triad signals. A bank of comb filters covering octaves 2 to 4 is utilised. The onsets of the guitar signals are located in the first frame. Then, the signals become more harmonic as they reach the steady state. The results are shown in Table 9-5.

| **Frame number** | **1** | **2** | **3** | **4** |
|---|---|---|---|---|
| **Correct detections** | 3/13= 23.07% | 6/13 = 46.15% | 8/13 = 61.53% | 10/13 = 76.9% |

**Table 9-5: Triad detection results**

**TEST 2: Detection of Guitar Triads and Tin whistle notes**

The tin whistle notes [CC, CD, CD#, CE, CF, CF#, CG, CG#, CA, CA#, CB] are mixed with different guitar major or minor triads. Each tin whistle note is mixed with a triad, whose root is in the same note played by the tin whistle (e.g., tin whistle and guitar play D5 and a D major triad respectively), and also with a triad whose notes do not feature in the tin whistle note (e.g., tin whistle and guitar play D5 and an E minor triad respectively). Thus, the database comprises 14 signals. A bank of comb filters covering octaves 5 to 6 is utilised to detect the tin whistle notes. The triad detection is performed as in TEST 1. The results are shown in Table 9-6.

| **Frame number** | **1** | **2** | **3** | **4** |
|---|---|---|---|---|
| **Correct guitar triads** | 4/14 = 28.57% | 8/14 = 57.14% | 9/14 = 64.28% | 8/14 = 57.14% |
| **Correct Tin whistle notes** | 12/14 = 85.71 % | 12/14 = 85.71 % | 13/14 = 92.85% | 9/14 = 64.28% |

**Table 9-6: % of correct detection of guitar triads and tin whistle notes**

## 9.4 Discussion

In Section 9.3, the performance of the MPECF has been evaluated for three different databases of audio signals. The first database comprises synthetic signals, the second one real guitar triads and the third one real guitar triads accompanying tin whistle signals.

Firstly, the MPECF is compared against the [Tadokoro '03] pitch detector, whose model forms the basis of the MPECF. In order to avoid octave pitch detection errors in the comparison, the comb filters and the audio signals are generated within octave 4. Thus, the comparison focuses on the comb filter utilised. The MPECF provides 100% of correct detection in all the tests performed, which includes monophonic signals (Table 9-2) and a range of two (Table 9-2), three (Table 9-3) and four (Table 9-4) note polyphonic mixtures. By using a frame length $L$ equal to 8192 samples, [Tadokoro '03] method correctly detects 100 % of the notes in the case of monophonic signals. However, in the case of using polyphonic signals the system detects correctly 90.9%, 83.3% and 85 % of the notes in the two, three and four notes polyphony respectively. As opposed to the MPECF, a shorter frame $L$ equal to 4096 samples degrades the accuracy of the detection using the [Tadokoro '03] method, with the percentage of correct detections in a 4 notes polyphony equal to 30 %. These results show the advantage of replacing the standard FIR comb filters by the comb filter described by Equation (96).

The MPECF is also evaluated by using real signals. Firstly, the system is evaluated for a database of 13 major and minor guitar triads. The results are shown in Table 9-5, where it can be seen that the system is capable of detecting 76.9 % of correct results when analysing the purely harmonic part of the guitar. However, the performance varies depending on the analysis frame, where the pitch detection analysis in the frame containing the guitar onset provides only 18.75 % of correct results. This difference in accuracy is due to the non-harmoncities that occur during the onset part of a musical signal, which affects the performance of the multi-pitch estimator.

Next, the MPECF is evaluated for a database of 14 audio signals containing mixtures of tin whistle notes with different guitar major or minor triads. The results are shown in Table 9-6, where it can be seen that the performance of the triad detector is not altered by the presence of the tin whistle. The tin whistle plays in a different octave range, and since the leading instrument is generally played louder than the harmonic accompaniment, the percentage of correct tin whistle note detection is high for the first three frames, with a percentage of correct detections in the 85.71-92.85 % range. Since some of the tin whistle signals utilised in the test release before the fourth frame in the analysis, the results degrade for this frame (64.25 % of correct detections).

## 9.5 Conclusions

A multi-pitch detection system, MPECF, based on an existing pitch detection model has been presented in this chapter [Tadokoro '03]. The MPECF improves the accuracy of this pitch detection model by utilising a different comb filter type, which combines a standard FIR and IIR comb filter. This filter provides a good compromise between the notch sharpness, the pass-band response flatness and the width of the filter null. In addition, in an effort to avoid low octave pitch detection errors, the magnitude response of the filter is weighted by using a method utilised by [Martin '82, Morgan '97] to detect the pitch of speech. A method that detects the harmonic triads provided by the accompaniment and the notes played by the leading instrument is also provided. The development of the complete system corresponds to Contribution 4 in Section 1.2.1.

The system has been evaluated using three different databases, comprising synthetic monophonic and polyphonic signals, real guitar triads, and mixtures of guitar triads accompanying tin whistle tunes. The results are accurate for all of the databases, where the MPECF system is capable of detecting four simultaneous notes in a polyphony (three note chords (triads) and a tin whistle note). However, it has also been shown that the performance of the MPECF in the real signals database depends on the position of the analysing frame relative to the instrument onset. This proves that the use of a robust onset detector is crucial in music transcription to detect the commencement of the note, as well as to avoid incorrect detections due to the inharmocities that arise during the onset part of a signal.

As described in the introduction of this chapter, Irish Traditional music is accompanied after knowing the mode of the tune [James '02, McQuaid '05]. In addition, major and minor triads are generally sufficient to provide the accompaniment [James '02]. As seen in Section 2.2 and 2.3, there are only a limited range of modes that are commonly utilised in Irish traditional music. In addition, Irish traditional music tunes generally end on the pitch of the mode (e.g., D in D Ionian) [Larsen '03]. By combining all this knowledge, the range of comb filter notes to consider in the detection by the MPECF can be reduced. In addition, by knowing the mode, errors caused by the MPECF can be corrected. The configurations of [Tadokoro '03] and MPECF models are adequate for this purpose, and can be customised to the mode notes. The development of a mode detector to investigate the improvement in the results can be considered as an area of future research.

The system assumes that the number of notes played together is known. However, in order to provide a fully automated transcription system, a method that detects the number of notes present in the polyphony can be considered.

Finally, the efficiency of the system can be improved upon by using polyphase filters, whose development might also be considered as an area of future work.

# 10 Summary and future work

This thesis presents different signal processing algorithms for the purpose of transcribing Irish traditional music. A significant feature of this music is ornamentation, which is an improvised expression of the style of the player. These techniques have been learned by listening and adaptation from other players, explaining the absence of transcriptions including ornamentation and the lack of consensus among players regarding ornamentation notation and types. Irish Traditional music players exploit the full potential of ornamentation in solo performances. In the case of group playing, Irish traditional music is generally played in unison. Nevertheless, simplistic harmonic accompaniment has also been incorporated in recent years.

A review of existing onset detection methods concluded that the main problems encountered by existing approaches are related to frequency and amplitude modulations, in fast passages such as legato, in the detection of slow onsets, and in detecting ornamentation events. A review of existing pitch detection methods was also undertaken, which highlights that a system that detects the different types of ornamentation within Irish traditional music has not yet been implemented. In addition, the review shows that periodicity based methods are less accurate in polyphonic signals.

In order to overcome the problems identified in the literature review, different applications for onset, pitch and ornamentation detection were presented in this thesis in Chapters 6 to 9, which are referred to in Section 1.2.1 as Contribution 1 to Contribution 4 respectively. A number of conclusions related to each of the contributions can be found in a section of each of these chapters (Sections 6.5, 7.5, 8.5 and 9.5).

Firstly, an onset detection method which focuses on the characteristics of the tin whistle within Irish traditional music was developed (ODTW, Contribution 1) [Gainza '04c]. The tin whistle is an important instrument in Irish traditional music which frequently produces amplitude and frequency modulations. The instrument has a legato nature of playing, represents a good example of a slow onset instrument and reflects well the use of ornamentation within Irish traditional music. Consequently, all the problems related to onset detection in the literature review are also encountered in the detection of the tin whistle onset. The onset detection system utilises knowledge of the notes and modes that the tin whistle is more likely to produce, and expected blowing pressure that a tin whistle produces per note. The problems associated to legato playing in onset detection are catered for by utilising a multi-band decomposition, where one band is utilised per note. In an effort to reduce the effect of amplitude modulations, different novel thresholding methods were implemented. By using these methods in conjunction with an optimisation of other system parameters, the onset detection system deals with moderate signal amplitude modulations. However, the system is not capable of dealing with strong amplitude and frequency modulations. In addition, the problem related to the detection of ornamentation events in onset detection systems is not overcome by the system, which assumes that close onset candidates belong to the same onset. The latter limitation is overcome by the ornamentation detector of Contribution 2. This system forms note and ornamentation candidate segments, which are treated separately. Following this, the segments are combined to form single and multi-note ornamentation, which are described by a set of rules concerning general ornamentation theory. This attempt to transcribe the most common types of ornamentation represents a novelty in the field, since it has never been attempted before. In addition, the onset time estimation provided by this system reflects well Irish traditional music features, since the onset is estimated at the beginning of the ornamentation event.

The problems of strong amplitude and frequency modulations are still present in the ornamentation detector. However, these limitations are overcome by Contribution 3, which represents a novel onset detection system (ODCF) based on the harmonicity of the signal. This signal property is captured by the use of a bank of FIR comb filters, which also has a harmonic magnitude response. In addition, the system overcomes the difficulties that existing approaches encounter when detecting slow onsets, providing a more accurate onset time than these approaches and the ODTW. Consequently, all the difficulties encountered by the existing onset detection approaches have been dealt with by the systems represented in Contribution 1 to Contribution 3.

For the case of unison playing, existing periodicity based pitch detection methods such as FIR comb filters might be utilised to transcribe the notes. However, with the inclusion of harmonic accompaniment the performance of these methods degrades. In an effort to detect the accompaniment triads, a multi-pitch detection system was implemented (MPECF, Contribution 4), which combines the structure of the multi-pitch detection model of [Tadokoro '03] with the use of a more accurate comb filter and the weighting method of [Martin '82, Morgan '97]. The system detects the harmonic triads provided by a guitar accompanying a tin whistle. The results also show that the system fails to provide an accurate estimation of the pitch of the notes during the onset part of the signal. This highlights the importance of onset detection in music transcription.

## 10.1 Future work

A number of suggestions for future work related to each of the contributions can be found in a section of each of these chapters (see Sections 6.5, 7.5, 8.5 and 9.5).

Ornamentation characteristics vary depending on the style of the player, which complicates the task of establishing a set of rules to describe this technique. Thus, the creation of a corpus of different tin whistle signals, played by different players and styles, warrants future work. The corpus may also be used by the ODTW to improve the accuracy of the detection by investigating the different lilt of tin whistle players.

The development of a mode detector has also been suggested as an area of future work in the ODTW and in the MPECF. A mode detector will customise the multi-band decomposition to the notes described by the mode, which increases the robustness of the onset detector in legato playing. In addition, the use of an instrument recognition system will also be utilised to configure the system to the characteristics of other instruments. The same mode detector is likely to improve the performance of the MPECF by limiting the amount of notes considered in the detection, and also correcting pitch detection errors. For example, the results can be compared against the chords typically played by each of the tune modes, which are provided in [James '02].

Finally, the efficiency of the comb filters utilised by the ODCF and the MPECF should improved upon by the use of polyphase filters.

# Appendix 1: Abbreviations

| | |
|---|---|
| *acc* | Detection Accuracy Estimation (defined in page 115 for onset detection and in page 148 for ornamentation detection) |
| *acf* | Autocorrelation function |
| *Corr* | Correct Ornamentation Detections (defined in page 148) |
| *Fal* | False Ornamentation Detections (defined in page 148) |
| *fc* | Filter cut-off frequency |
| *f0* | Fundamental frequency |
| *fs* | Sampling Frequency |
| HFC | High Frequency Content |
| *FN* | False negative Onsets (defined in page 115) |
| *FP* | False positive Onsets (defined in page 115) |
| *lc* | Long crann (ornamentation type) |
| *lr* | Long roll (ornamentation type) |
| LPF | Low Pass Filter |
| MPECF | Multi-pitch Estimator based on Comb Filters |
| ODCF | Onset Detection System based on Comb Filters |
| ODTW | Onset Detection System applied to the Tin Whistle |
| OrnTr | Ornamentation Transcription |
| *Orn* | Ornamentation |
| *pFN* | Percentage of False Negative Onsets (defined in page 117) |
| *pFP* | Percentage of False Positive Onsets (defined in page 117) |

| | |
|---|---|
| *pGP* | Percentage of Good Positive Onsets (defined in page 117) |
| *pCorr* | Percentage of Correct Ornamentation Detections (defined in page 148) |
| SACF | Summary Autocorrelation function |
| *sc* | Short crann (ornamentation type) |
| *Sg* | Audio Segment |
| *sr* | Short roll (ornamentation type) |
| STFT | Short Time Fourier Transform |
| *Str* | Strike (ornamentation type) |
| *TM* | Thresholding Method (defined in page 115) |
| *win* | Sliding window utilised in onset detection (defined in page 116) |

# Appendix 2: Contents of the Companion CD

A summarised description of the contents of the companion CD is provided in this Appendix, which includes the complete code of the contributions implemented in this thesis and the database of signals utilised to perform the different tests.

## 1. MATLAB Code

The code has been generated by utilising the technical computing software MATLAB [MATLAB '06]. The m-files are organised by application name (ODTW, OrnTr, ODCF and MPECF). As an example, all the files corresponding to the application ODTW are included in a folder called m-files\ODTW. A complete list of the m-files is provided as follows:

- **m-files\ODTW**

**main.m**: provides an example of the use of the ODTW.

**onsetDec.m**: main m-file to call. Detects tin whistle onsets.

**getEn.m**: obtains the energy envelopes of the bands.

**notesFreqsAndBins.m**: note frequency-bin conversion.

**makeEnarray.m**: performs a multi-band decomposition.

**smoothing.m:** smooths the energy envelopes.

**noNote.m**: obtains the active bands that contain notes.

**setThreh.m**: sets the band thresholds.

**getPeaks.m**: obtains the peaks, which correspond to the band onset candidates.

**peakPicking.m**: picks the onset candidates that fulfil certain criteria.

**compRes.m:** compares the detected onset locations against the real onset locations.

- **m-files\OrnTr**

**main.m**: provides an example of the use of the OrnTr.

**ornDec.m**: main m-file to call. Detects single and multi-note ornamentation.

**getEn.m**: id. at ODTW.

**notesFreqsAndBins.m**: id. at ODTW.

**makeEnarray.m**: id. at ODTW.

**smoothing.m:** id. at ODTW.

**noNote.m**: id. at ODTW.

**setThreh.m**: id. at ODTW.

**makeSeg.m**: forms note and ornamentation segments.

**ornTrans.m:** transcribes single-note ornamentation.

**Cut.m**: transcribes cuts.

**missStr.m:** investigates the occurrence of undetected strikes in repeating notes.

**MultiOrnTrans.m:** transcribes multi-note ornamentation.

**linkSeg.m:** fill frames that have not been detected as part of notes or ornamentation .events with untranscribed ornamentation segment candidates.

**getPeaks.m**: id. at ODTW.

**peakPicking.m**: id. at ODTW.

**compResOrn.m:** compares the detected ornamentation events against the real ornamentation locations.

- **m-files\ODCF**

**main.m**: provides an example of the use of the ODCF.

**ODCF.m**: main m-file to call. Generates an onset detection function based on a comb filters technique.

**CFilter.m**: calculates the spectral fit of the comb filters spectra with the signal spectrum.

**isNote.m**: detects the frames where no note was playing.

**OffOnsetTran.m**: detects the slow offset-onset transition and cancels the onset candidate produced by the occurrence of an offset.

**getPeaks.m**: id. at ODTW.

- **m-files\MPECF**

**main.m**: provides an example of the use of the MPECF.

**MPE.m**: main m-file to call. Multi-pitch estimator based on comb filters.

**triadTrans.m**: transcribes the triads that the detected notes by MPE.m form.

## 2. Signal Databases

In addition, the database of signals utilised in order to evaluate the performance of the algorithms are also provided. The different databases are included in a folder called "signal databases", and are listed as follows:

- **signal databases\Tin whistle signals database**: database of Irish tin whistle signals utilised by the ODTW, ODCF and OrnTr. This database also includes the real location of the onset and ornamentation events, which have been manually labelled.

- **signal databases\Irish trasditional instrument signals database**: database of Irish traditional music instruments utilised by the ODCF. This database also includes the real location of the onsets, which have been manually labelled.
- **signal databases\MPECF signal database**: this database includes major and minor guitar triad solos and mixtures of guitar triads accompanying tin whistle notes. This database is utilised by the MPECF.

# References

[Amatriain '02] Amatriain, X. et al., 2002. *Spectral Processing*. In *Digital Audio Effects, DAFX*. 1st ed, John Wiley & Sons, pp. 373-438.

[Asari '04] Asari, H., 2004. "Non-negative Marix Factorization: a possible way to learn sound dictionaries". Technical Report

[Beardah '95] Beardah, C. C. "The Kernel Density Estimation toolbox for Matlab". Available from: http://euler.ntu.ac.uk/maths.html. Accessed 2004.

[Bello '03] Bello, J. P. and Sandler, M., 2003. "Phase-based note onset detection for music signals". In Proc of *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)*.

[Bello '04] Bello, J. P. et al., 2004. "On the use of phase and energy for musical onset detection in the complex domain". *Signal Processing Letters, IEEE*, vol. 11 (6), pp. 553 - 556.

[Bernhardt '74] Bernhardt, P. A., 1974. "Separation of Lunar and Solar Periodic Effects in Data". *Journal of geophysical research*, vol. 79 (28), pp. 4343-4349.

[Bilmes '93] Bilmes, J. A., 1993. *Timing is of the Essence: Perceptual and Computational Techniques for Representing, Learning, and Reproducing Expressive Timing in Percussive Rhythm*. MSc Thesis. MIT.

[Bregman '90] Bregman, A., 1990. *Auditory Scene Analysis*. MIT Press.

[Brother '06] Brother, S. "Brother Steve's tin-whistle pages". Available from: http://www.rogermillington.com/siamsa/brosteve/index.html. Accessed May 2006.

[Brown '92] Brown, J. C., 1992. "Musical fundamental frequency tracking using a pattern recognition method". *Journal of the Acoustical Society of America*, vol. 92 (3), pp. 1394-1402.

[Brown '93] Brown, J. C., 1993. "Determination of the meter of musical scores by autocorrelation". *Journal of the Acoustical Society of America*, vol. 4 (94), pp. 1953-1957.

[Bulmer '79] Bulmer, M. G., 1979. *Principles of Statistics*. Courier Dover Publications.

[Carolan '06] Carolan, N. "Irish Traditional Music Archive. ITMA. What is Irish Traditional Music? (Definition and characteristics)". Available from: http://www.itma.ie/home/leaf1a.htm. Accessed January 2006.

[Carson '99] Carson, C., 1999. *Irish Traditional music*. Appletree Press (UK).

[Chafe '86] Chafe et al., 1986. "Source Separation and Note Identification in Polyphonic Music". In Proc of *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '86)*.

[Cool '02] Cool Edit Pro 2.00, 2002. Syntrillium Software Corporation.

[De Cheveigne '91a] De Cheveigne, A., 1991a. "Speech f0 extraction based on Licklider's pitch perception model". In Proc of *International Congress in Phonetic sciences, ICPhS*.

[De Cheveigne '91b] De Cheveigne, A., 1991b. "A Mixed Speech F0 Estimation Algorithm". In Proc of *Eurospeech Conference*.

[De Cheveigne '02] De Cheveigne, A. and Kawahara, H., 2002. "YIN, a fundamental frequency estimator for speech and music". *J. Acoust. Soc. Am.* pp. 1917-1930.

[Dolson '86] Dolson, M., 1986. "The Phase Vocoder: A tutorial". *Computer Music Journal*, vol. 10 (4).

[Dorran '04] Dorran, D. and Lawlor, R., 2004. "Time-scale modification of music using a synchronized subband/time-domain approach". In Proc of *IEEE International Conference on Acoustics, Speech and Signal Processing*. Montreal.

[Duggan '06a] Duggan, B., 2006a, Personal Communication.

[Duggan '06b] Duggan, B., Zheng, C. and Cunningham, P., 2006b. "MATT - A System for Modelling Creativity in Traditional Irish Flute Playing". In Proc of *ECAI'06 workshop MT2*. Riva del Garda, Italy.

[Dutilleux '98] Dutilleux, P., 1998. "Filters, Delays, Modulations and Demodulations:A Tutorial". In Proc of *Digital Audio Effects, DAFX.*

[Duxbury '01] Duxbury, C., Davies, M. and Sandler, M., 2001. "Separation of transient information in musical audio using multiresolution analysis techniques". In Proc of *Digital Audio Effects (DAFX)*. Limerick, Ireland.

[Duxbury '02] Duxbury, C., Davies, M. and Sandler, M., 2002. "A hybrid approach to musical note onset detection". In Proc of *5th Int. Conference on Digital Audio Effects (DAFx-02)*. Hamburg, Germany.

[Duxbury '03a] Duxbury, C. et al., 2003a. "A combined phase and amplitude based approach to onset detection for audio segmentation". In Proc of *4th European workshop on Image Analysis for Multimedia Interactive Services (WIAMIS-03)*. London (UK).

[Duxbury '03b] Duxbury, C. et al., 2003b. "Complex Domain Onset Detection For Musical SIgnals". In Proc of *6th Int. Conference on Digital Audio Effects (DAFx-03)*. London, UK.

[Duxbury '04] Duxbury, C. et al., 2004. "A Comparison Between Fixed And Multiresolution Analysis For Onset Detection In Musical Signals". In Proc of *7th Int. Conference on Digital Audio Effects (DAFx'04)*. Naples, Italy.

[Ellis '96] Ellis, D., 1996. *Prediction-driven computational auditory scene analysis.* PhD Thesis. Massachusetts Institute of Technology.

[Eronen '00] Eronen, A. and Klapuri, A., 2000. "Musical Instrument Recognition Using Cepstral Coefficients and Temporal Features". In Proc of *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP.*

[Fernández-Cid '98] Fernández-Cid, P. and Casajús-Quirós, F. J., 1998. "Enhanced Quality and Variety for Chorus/Flange Units". In Proc of *Digital Audio Effects, DAFX.*

[Fletcher '98] Fletcher, N. H. and Rossing, T. D., 1998. *The Physics of Musical Instruments*. Springer.

[Gainza '04a] Gainza, M., Lawlor, B. and Coyle, E., 2004a. "Single-Note Ornaments Transcription For The Irish Tin Whistle Based On Onset Detection". In Proc of *7th Int. Conference on Digital Audio Effects (DAFX-04)*. Naples, Italy.

[Gainza '04b] Gainza, M., Lawlor, B. and Coyle, E., 2004b. "Harmonic Sound Source Separation using FIR Comb Filters". In Proc of *117th AES Convention*. San Francisco.

[Gainza '04c] Gainza, M. et al., 2004c. "Onset Detection and Music Transcription for the Irish Tin Whistle". In Proc of *Irish Signals and Systems Conference, ISSC.* Belfast.

[Gainza '05a] Gainza, M., Lawlor, B. and Coyle, E., 2005a. "Multi pitch estimation by using modified IIR Comb Filters". In Proc of *47th International Symposium focused on Multimedia Systems and Applications (ELMAR)*. Zadar, Croatia.

[Gainza '05b] Gainza, M., Lawlor, B. and Coyle, E., 2005b. "Onset Detection Using Comb Filters". In Proc of *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*.

[García '99] García, L. and Casajus-Quirós, J., 1999. "Separation of Musical Instruments Based on Perceptual and Statistical Principles". In Proc of *Digital Audio Effects, DAFx*. Trondheim, Norway.

[Godsmark '99] Godsmark, D. and Brown, G., 1999. "A blackboard architecture for computational auditory scene analysis". *Speech Communication*, vol. 27 (3-4), pp. 351–366.

[Grey '75] Grey, J. M., 1975. *An exploration of musical timbre using computer-based techniques for analysis*. Ph.D Thesis.

[Howard '01] Howard, D. and Angus, J., 2001. *Acoustics and Psychoacoustics*. 2nd ed., Focal Press.

[Hoyer '02] Hoyer, P. O., 2002. "Non-negative sparse coding". In Proc of *12th IEEE Workshop on Neural Networks for Signal Processing*.

[James '02] James, D. "Guide to accompaniment for Irish traditional music". Available from: http://www.tiompanalley.com/index_files/tunes/accompan.htm. Accessed March 06.

[Kaiser '77] Kaiser, J. and Hamming, R., 1977. "Sharpening the response of a symmetric nonrecursive filter by multiple use of the same filter". *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 25 (5), pp. 415-422.

[Kashino '95a] Kashino, K. et al., 1995a. "Organization of hierarchical perceptual sounds: Music scene analysis with autonomous processing modules and a quantitative information integration mechanism". In Proc of *Computational auditory scene analysis workshop, international joint conferences on artificial intelligence*.

[Kashino '95b] Kashino, K. et al., 1995b. "Application of Bayesian probability network to music scene analysis". In Proc of *Computational auditory scene analysis workshop, international joint conferences on artificial intelligence*.

[Klapuri '98] Klapuri, A., 1998. *Automatic Transcription of Music*. MSc Thesis. Tampere University of Technology.

[Klapuri '99] Klapuri, A., 1999. "Sound onset detection by applying psychoacoustic knowledge". In Proc of *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP*.

[Klapuri '01] Klapuri, A. and Virtanen, T., 2001. "Automatic Transcription of Musical Recordings". In Proc of *Consistent & Reliable Acoustic Cues Workshop, CRAC-01*. Aalborg, Denmark.

[Klapuri '03 ] Klapuri, A. et al., 2003 "Automatic transcription of music". In Proc of *Stockholm Music Acoustics Conference (SMAC 03)*. Stockholm, Sweden.

[Klapuri '04] Klapuri, A., 2004. *Signal Processing Methods for the Automatic Transcription of Music*. PhD Thesis.

[Klapuri '03] Klapuri, A. P., 2003. "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness". *IEEE Transactions on Speech and Audio Processing*, vol. 11 (6), pp. 804-816.

[Kunieda '96] Kunieda, N., Shimamura, T. and Suzuki, J., 1996. "Robust method of measurement of fundamental frequency by ACLOS: autocorrelation of log spectrum". In Proc of *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'96.*

[Lahat '87] Lahat, M., Niederjohn, R. and Krubsack, D., 1987. "A spectral autocorrelation method for measurement of the fundamental frequency of noise-corrupted speech". *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35 (6), pp. 741-750.

[Larsen '03] Larsen, G., 2003. *The Essential Guide to Irish Flute and Tin Whistle*. Mel Bay Publications.

[Lee '01] Lee, D. and Seung, H. S., 2001. "Algorithms for Non-negative Matrix Factorization". *Advanced Neural Information Processing Systems, NIPS*, vol. 13 pp. 556-562.

[Lindley '06] Lindley, M. 2006, "Equal temperament". Grove Music Online. http://www.grovemusic.com. Ed. L. Macy.

[M. Neal '96] M. Neal, R., 1996. *Bayesian Learning for Neural Networks*. Springer.

[Marolt '02] Marolt, M. et al., 2002. "On detecting note onsets in piano music". In Proc of *11th Mediterranean Electrotechnical Conference. MELECON.*

[Martin '94] Martin, J., 1994. *The Acoustics of the Recorder*. Moeck.

[Martin '96a] Martin, K., 1996a. "Automatic transcription of simple polyphonic music: Robust front end processing." MIT Media Laboratory, Technical Report 399.

[Martin '96b] Martin, K., 1996b. "A blackboard system for automatic transcription of simple polyphonic music". MIT Media Laboratory, Technical Report 385.

[Martin '82] Martin, P., 1982. "Comparison of pitch detection by cepstrum and spectral comb analysis". *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '82*, vol. 7 pp. 180- 183.

[Masri '96a] Masri, P., 1996a. *Computer Modeling of Sound for Transformation and Synthesis of Musical Signal*. Ph.D Thesis. University of Bristol.

[Masri '96b] Masri, P. and Bateman, A., 1996b. "Improved Modelling of Attack Transients in Music Analysis-Resynthesis". In Proc of *International Computer Music Conference (ICMC).*

[MATLAB '06] MATLAB, 2006. The MathWorks, Inc.

[Mc Cullough '87] Mc Cullough, L. E., 1987. *The Complete Tin Whistle Tutor*. New York Oak Publications.

[McQuaid '05] McQuaid, S., 2005. *The Irish DADGAD Guitar Book*. Music Sales.

[Meddis '91a] Meddis, R. and Hewitt, M. J., 1991a. "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Phase sensivity". *The Journal of the Acoustical Society of America*, vol. 89 (6), pp. 2883-2894.

[Meddis '91b] Meddis, R. and Hewitt, M. J., 1991b. "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification". *The Journal of the Acoustical Society of America*, vol. 89 (6), pp. 2866-2882.

[Meddis '92] Meddis, R. and Hewitt, M. J., 1992. "Modeling the identification of concurrent vowels with different fundamental frequencies". *The Journal of the Acoustical Society of America*, vol. 91 (1), pp. 233-245.

[Miwa '99a] Miwa, T. and Tadokoro, Y., 1999a. "Musical Pitch Estimation of Different Instrument Sounds using comb filters for transcription". In Proc of *Midwest Symposium on Circuits and Systems.*

[Miwa '99b] Miwa, T., Tadokoro, Y. and Saito, T., 1999b. "Musical pitch estimation and discrimination of musical instruments using comb filters for transcription". In Proc of *42nd Midwest Symposium on Circuits and Systems.*

[Miwa '00] Miwa, T., Tadokoro, Y. and Saito, T., 2000. "The Problems of Transcription using Comb Filters for Musical Instrument Sounds and Their Solutions". Technical Report of IEICE SP2000-59.

[Moore '97] Moore, B. C. J., Glasberg, B. R. and Baer, T., 1997. "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness". *J. Audio Eng. Soc.*, vol. 45 (4), pp. 224-240.

[Moorer '74] Moorer, J. A., 1974. "The optimum comb method of pitch period analysis of continuous digitized speech". *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 22 (5), pp. 330- 338.

[Moorer '75] Moorer, J. A., 1975. *On the segmentation and analysis of continuous musical sound by digital computer*. PhD Thesis. Stanford University.

[Moorer '85] Moorer, J. A., 1985. "About this Reverberation Business". *Computer Music Journal*, vol. 3 (2), pp. 13-28.

[Morgan '97] Morgan, D. P. et al., 1997. "Cochannel speaker separation by harmonic enhancement and suppression". *IEEE Transactions on Speech and Audio Processing*, vol. 5 (5), pp. 407-424.

[Nehorai '86] Nehorai, A. and Porat, B., 1986. "Adaptive comb filtering for harmonic signal enhancement". *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34 (5), pp. 1124 - 1138.

[Noll '64] Noll, A. M., 1964. "Short -Time Spectrum and "Cepstrum" Techniques for Vocal-Pitch Detection". *The Journal of the Acoustical Society of America*, vol. 36 (2), pp. 296-302.

[O'Canainn '93] O'Canainn, T., 1993. *Traditional music in Ireland.* Music Sales Corporation.

[Pal '05] Pal, N. and Sarkar, S., 2005. *Statistics: Concepts and Applications*. Prentice Hall of India.

[Proakis '95] Proakis, J. G. and Manolakis, D. G., 1995. *Digital Signal Processing: Principles, Algorithms and Applications*. 3rd ed., Prentice Hall.

[Prout '06] Prout, E. and Donington, R. 2006, "All'unisono". Grove Music Online. http://www.grovemusic.com. Ed. L. Macy.

[Ross '74] Ross, M. et al., 1974. "Average magnitude difference function pitch extractor". *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 22 (5), pp. 353 - 362.

[Scheirer '98] Scheirer, E., 1998. "Tempo and Beat Analysis of Acoustic Musical Signals". *J. Acoust. Soc. Am.*, vol. 103 (1), pp. 588-601.

[Serra '89] Serra, X., 1989. *A system for sound analysis/transformation/synthesis based on a deterministic plus stochastic decomposition.* Ph.D Thesis. Stanford University.

[Settel '94] Settel, Z. and Lippe, C., 1994. "Real-time Musical Applications using FFT-based Resynthesis". In Proc of *International Computer Music Conference.* Aarhus.

[Slaney '93] Slaney, M., 1993. "An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank". Apple Technical Report

[Smaragdis '03] Smaragdis, P. and Brown, J. C., 2003. "Non-negative matrix factorization for polyphonic music transcription". In Proc of *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics.*

[Smith '97] Smith, S., 1997. *The Scientist and Engineer's Guide to Digital Signal Processing.* California Technical Publishing

[T. Jolliffe '02] T. Jolliffe, I., 2002. *Principal Component Analysis.* Springer.

[Tadokoro '01] Tadokoro, Y. and Yamaguchi, M., 2001. "Pitch detection of duet song using double comb filters". In Proc of *ECCTD.*

[Tadokoro '02] Tadokoro, Y., Matsumoto, W. and Yamaguchi, M., 2002. "Pitch Detection of Musical Sounds using Adapative Comb filters controlled by Time Delay". In Proc of *IEEE International Conference on Multimedia and Expo, ICME '02.*

[Tadokoro '03] Tadokoro, Y., Morita, T. and Yamaguchi, M., 2003. "Pitch detection of musical sounds noticing minimum output of parallel connected comb filters". In Proc of *Conference on Convergent Technologies for Asia-Pacific Region, TENCON.*

[Tolonen '00] Tolonen, T. and Karjalainen, M., 2000. "A computationally efficient multipitch analysis model,". *IEEE Trans. Speech Audio Processing*, vol. 8 (6), pp. 708-716.

[Vallely '99] Vallely, F., 1999. *The companion to Irish Traditional Music.* New York University Press.

[Virtanen '00] Virtanen, T. and Klapuri, A., 2000. "Separation of Harmonic Sound Sources Using Sinusoidal Modeling". In Proc of *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP.*

[Wise '76] Wise, J., Caprio, J. and Parks, T., 1976. "Maximum likelihood pitch estimation". *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 24 (5), pp. 418- 423.