Musically Significant, Automatic Localisation of Note Boundaries for the Performance Analysis of Vocal Music.

Nicholas J Bailey, Douglas McGilvray, Graham Hair

May 13, 2008

1 Introduction

Hair's recent compositions for voice include works composed in the 19-ET system. One of the primary motivations for the move away from 12-ET is a more extensive chromatic variety [11]. Works such as Ash require a significant effort to "retune the ear" of conservatory-trained musicians: they have spent most, if not all, of their careers performing in standard 12-ET tuning or at least in tuning systems which explicitly admit only twelve divisions of the octave.

Having delivering advanced, 19-ET capable ear-training software, there is a motivation better to understand the emergent performance structures, essentially the performance tradition, of such music. Previous published work on automatic transcription determines note onset simply by awaiting a temporarily stable pitch, or removes the effect of vibrato by averaging the pitch trajectory over time. A detailed measurement of the pitch trajectories of expert performers, and separately, of each note's inflections needs to be made, so that this information can be interpreted in the context of composed harmonic and melodic structures.

We describe two specially developed segmentation algorithms in application to vocal music. In both cases, particular attributes of the music would make it likey that naïve algorithms would fail. Firstly, pitch-based segmentation is undertaken in a piece where the smallest chromatic interval is 1/19th of an octave. This presents the algorithm with pitch resolution problems more stringent than pitch-based segmentation in 12ET. The algorithm has been deployed as a rehearsal aid for repertoire-specific training of expert musicians at the Royal College of Music[5]. The second example demonstrates the use of vector quantisation to establish significant pitch-trajectory characteristics in a recording of Schoenberg's Pierrot Lunaire. The source material was recorded especially by the soprano Jane Manning, an internationally recognised authority on the performance of this piece whose first recording of it dates from 1967[9].

2 A Vocal Segmentation Algorithm for Discrete-Pitch Music

There are a plethora of transcription algorithms available in academia and commercially. Typically these transcription algorithms can be described as three processes: feature analysis, segmentation and quantisation [2, 12]. The algorithm presented here is a *segmentation* algorithm rather than a transcription algorithm. In other words, the algorithm is designed to segment a performance according to certain criteria. This algorithm is intended to be used to identify the regions of stationary pitch within performed notes, for the purposes of analysing pitch accuracy. Although this is similar in many respects to the first two stages of automatic segmentation, there are very important differences. This application does not require an exact onset which corresponds to the beginning of the note (whether that refers to the perceptual onset or the physical onset). Notes consist of unpitched sections, often within the attack of the note which should be considered when identifying the onset of a note. However, these sections are disregarded when identifying notes solely on the basis of frequency.

Additionally, within the context of microtonal rehearsal, many of the reliable cues used for automatic transcription, or other types of segmentation algorithm are missing. The two most important cues used in transcription are pitch and dynamics. In vocal performance, it is very common for this dynamic information to be absent from note transitions. Performances are often sung legato where the transition between notes is only indicated by a change in pitch rather than dynamics. Therefore, an application which aims to transcribe vocal performance within a rehearsal environment must be capable of detecting transitions between notes solely on the basis of pitch information.

Unfortunately, in the context of microtonal rehearsal, pitch is also not as reliable as it is in the transcription of standard repertoire. Even, and perhaps one should say especially, expert performers very often approach this repertoire with little experience of the scales involved, and have to retrain themselves to the new scales and harmonic contexts. This is particularly true of works such as the examples here which call for ensemble performance. A useful rehearsal tool has to be sufficiently robust to accommodate the pitch of a note being very inaccurate, and possibly very unstable. There are two common methods of identifying note transitions using only pitch information. The first, rather simple method, identifies transitions between notes at the boundaries between different pitches. This method effectively quantises the trajectory of frequency values into a sequence of pitch values, or scale intervals. This is often used in simple monophonic transcription routines such as those used to convert audio files into MIDI data.

The algorithm of Weihs and Ligges [17] uses such a method. The frequency trajectory is initially quantised to obtain a sequence of pitches. This 'pitch' trajectory is then smoothed to suppress vibrato. The resulting list of pitch values is then segmented and quantised in time to produce a transcription. This method is only of use in the segmentation of accurate performances. For example, if the frequency trajectory of a note crosses a pitch boundary, the note will be segmented into two separate notes rather than one (extremely out-of-tune) note. Furthermore, it discards the actual frequency information which is necessary in the analysis of pitch accuracy, or intonation.

The second method for detecting note transitions based on pitch uses the pitch gradient as a cue. However, vocal performance often contains vibrato which results in continuous oscillation in the derivative of the frequency trajectory. This vibrato must be suppressed to detect transitions between notes. Rossignol et al [13] describe several methods for the detection and quantification of vibrato in musical performance based on spectral modelling, spectral envelope distortion, auto regressive prediction and a method based on the analysis of maxima and minima in the vibrato. Those based on the spectrum of the signal, rather than the frequency trajectory, will be affected by the presence of other instruments in the recording, whereas those based on the frequency trajectory can be applied to the output of any transcription algorithm. The maxima-minima method is among the most promising, though it has yet to be tested in the current application. This method first detects the maxima and minima of the frequency trajectory. Interpolation is then used to determine the two trajectories which follow the maxima and minima points independently. This method allows properties of the vibrato to be analysed such as frequency and amplitude. The geometric mean of the two calculated trajectories also predicts a frequency trajectory for the performance with the vibrato suppressed. This was used in a system for feature extraction and acoustic segmentation [14]. This system combined a speech/singing/noise discriminator optionally followed by the aforementioned vibrato suppression (for musical segmentation) and a multi-agent segmentation algorithm. The segmentation algorithm comprised algorithms analysing nine features including derivatives of frequency trajectory and energy, inharmonicity, a voicing coefficient, probabilistic pitch transition and AR modelling.

McNab et al [12] presented an automatic transcription system developed and used in applications requiring vocal transcription. The system describes an algorithm which which advances a window of analysis through the frequency trajectory in order to segment it. When a segment has been found of length 100ms in which all frequency estimates are within 50 cents of the mean frequency of the window, the segment is assumed to correspond to a note. The boundaries of the segment may then be extended as long as the previous criteria holds true. In many ways, this algorithm could be easily adopted to microtonal use. The threshold of $\pm$ 50 cents relates to the smallest interval in 12ET. This could be adjusted for each microtonal tuning. Also, the threshold is based on the mean frequency of the segment, which means that the algorithm is to some degree less affected by notes which are performed out of tune. In fact, the paper mentions the application of the system to transcribing just and Pythagorean tuning, though there is no mention of any changes to the thresholds for that purpose. However, the algorithm seems to have little accuracy in identifying steady portions of a note. Indeed, the algorithm identifies a glissandi as a series of segments which is contrary to the requirements of both microtonal rehearsal and automatic transcription.

2.1 Method

The piecewise linear segmentation algorithm can be subdivided into three stages:

1. Fundamental Frequency Estimation
2. Preliminary Onset & Endpoint Detection

3. Onset & Offset Localisation

2.1.1 Fundamental Frequency Estimation

The pitch detection stage analyses the audio stream producing a list of frequency estimates. The pitch detection algorithm is a two stage algorithm based on the autocorrelation method which, during the use of the Rosegarden Codicil, was found to give more stable results for the analysis of soprano voice over other simple methods such as the Harmonic Product Spectrum (HPS) [16]. The autocorrelation is calculated using the Weiner Kinchen theorem (below), which for real-valued functions allows the auto-correlation to be calculated using the Fast Fourier Transform (FFT). It states that the autocorrelation of the time-domain signal is equal to the Fourier Transform of the power spectrum $S(f)$ of the signal. The initial estimate of fundamental frequency is estimated at the maximum value in the autocorrelation function after the initial peak at $\tau = 0$.

$$R(\tau) = \int_{\infty}^{-\infty} S(f)e^{j2\pi\tau}\, df$$

The second stage performs a localisation of the fundamental frequency by analysing the phase difference between successive FFT frames. Assuming that there is only one partial per FFT bin, its frequency can be calculated more precisely than the resolution of the FFT[4]. The change of phase of the partial from successive overlapping frames is an indication of the difference between a bin's centre frequency and the partial's. The calculation of the autocorrelation in the first stage means that the FFT data is already available from previous calculations. When there is no distinguishable peak, the frame is labelled as un-pitched.

2.1.2 Initial onset & endpoint detection

The second stage searches the pitch trajectory to determine initial estimates of note onsets and offsets. The frequency trajectory is first converted into cents to remove non-linearity. The algorithm then proceeds to search for onsets & offsets based on gradient. To avoid the false detection of offset and onsets, any vibrato must be suppressed. The algorithm iterates through the pitch trajectory calculating for each analysis window, an average gradient. The average gradient was calculated using a method of linear regression. This provides a 'line-of-best-fit' (LoBF) which represents an average linear trajectory which minimises the error using the least squares method between a point and the closest point to it on the line for all points in the analysis window.

The frequency trajectory was analysed with a window step of 256 samples and a sampling rate of 44100Hz. The onsets and endpoints were detected based on a threshold gradient of 77.5 cents per second analysing the frequency trajectory with a line which spanned 100 frequency estimates. These values were also chosen through qualitative assessment, using performances of the 19ET pieces composed by Professor Graham Hair. A purely scientific examination of how pitch gradient relates to how humans segment music would require an in-depth psychological study. This should also include an investigation of the effect microtonal tunings have on pitch segmentation. This is beyond the scope of this project, therefore a qualitative assessment must suffice until such work has been carried out.

When the gradient of the LoBF fell below the threshold, an onset was recorded at the *beginning* of the analysis frame. Conversely, when the gradient rose above the threshold, an offset was recorded at the *end* of that frame. This ensured that the recorded onset was consistently judged to be before the apparent onset of a note, and the recorded offset was consistently judged to be later than that of the actual offset.

2.1.3 Localisation of note boundaries

To localise the onset and endpoint points of each note, note candidates are created based on the initial onset and endpoints discovered in the previous stage. The localisation process also relies on linear regression and each note candidate is represented by a LoBF which is calculated based on each point in the pitch trajectory which lies within the onset and endpoint points. Localisation is an iterative procedure which followed the steps below.

1. The error for a candidate note is calculated between a frequency estimate and the closest point on the candidate line to that estimate. The error is calculated using the least squares method.

2. The LoBF and subsequent error is calculated for the case where the last point is removed from the original set.

3. The LoBF and subsequent error is calculated for the case where the first point is removed from the original set.

4. If the removal of neither of these two points result in a decrease in the error of the line, the process is halted. Otherwise, the point which caused the greatest decrease in error is removed.

5. This process is repeated until the potential decrease in error, as a fraction of the current error, is less than an arbitrary limit.

It is essential in the above procedure that the boundaries of the initial candidate are 'overestimated'. Due to the periodic nature of vibrato, there will be numerous sets of boundaries at which the error will exhibit a local minima. Therefore, the localisation must be performed using regression rather than extrapolation.

Figure 1 demonstrates the results of localisation. The diagrams show the analysis of the second phrase of *Ash* by Graham Hair which extends from bar 4 to the final A in bar 5 (see Figure 2). The initial candidate for each note is represented by a blue line and the final candidate after regression is indicated by the green line. Each unit on the vertical axis represents a frequency change of 20Hz; the horizontal axis is labeled in milliseconds from the start of the recording.

In some instances, two notes were incorrectly analysed as one continuous note. This occurred only in instances where the notes are just one hyperchromatic step apart, and the transition between the note is unclear. In these instances, the trajectory of the individual notes tends towards the contour of the phrase at that point. For example, Figure 4 shows two notes which have been

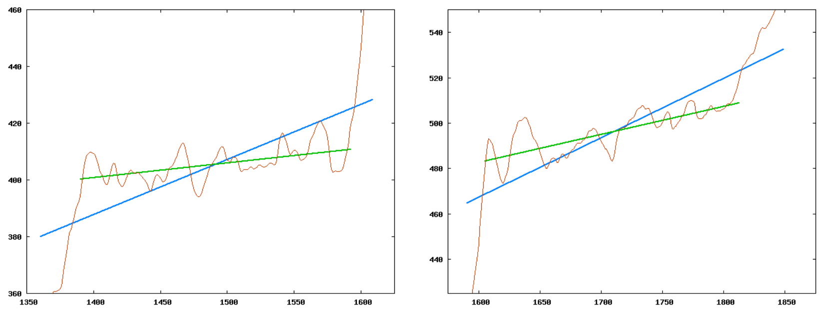

Figure 1: Rough (blue) & final (green) estimate of the 7th and 8th note of the extract from *Ash* using the pitch trajectory.

Figure 2: Bars 4 & 5 of *Ash* composed by Graham Hair. The second phrase of the piece is indicated.

incorrectly identified as a single note. The trajectory of the notes is steadily rising which obscures the transition by reducing its size. These two notes occur in bar 6 (Figure 3) within a run of five notes each one hyperchromatic step above the other. Repeated rising or falling hyperchromatic steps often proved to be the most difficult passages for the performers. Also, in tunings with more divisions of the octave, the transition between notes will obviously be less clear than those with fewer divisions of the octave. Therefore, the threshold could be adjusted according to the tuning to achieve optimal performance. Tunings with larger intervals could afford a more liberal threshold, whereas tunings with smaller intervals will require a tighter threshold.

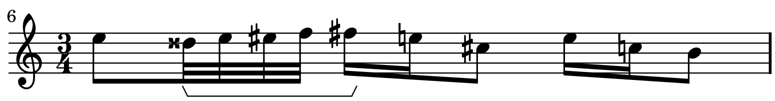

Figure 3: Bar 6 of *Ash* composed by Graham Hair. The bracket indicates the consecutive rising chromatic steps.

To reduce the number of incorrectly merged notes, notes whose trajectory spanned more than a threshold were split into two separate notes. The notes were split by finding the point within the original note's duration where the magnitude of the gradient was highest. At this point the line was split into two lines, and each line was regressed independently.

The ranges of some individual notes was found to be greater than that of some of the merged notes. Therefore, choosing the optimal threshold was a compromise between correcting the most errors and breaking the fewest correct identifications. In the course of this project, this threshold was found to be 0.8

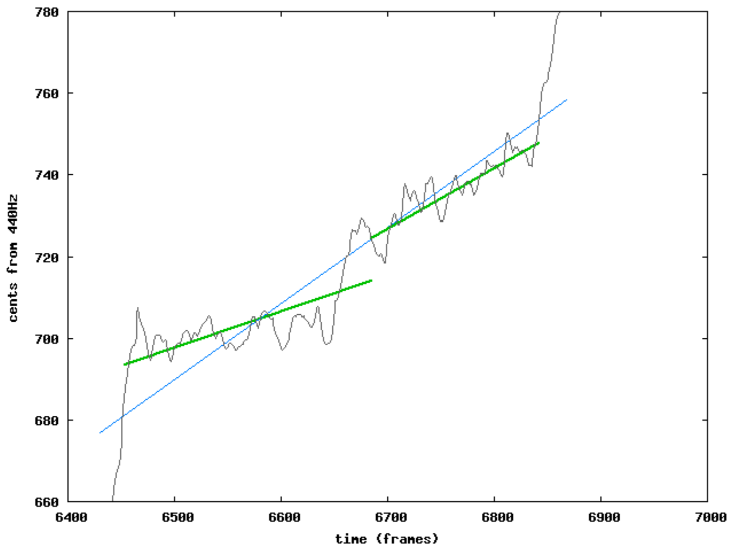

Figure 4: An example of the note splitting process. The blue line represents the original note candidate. The two green lines represent the two new candidates following the splitting process.

of a 19ET semitone. However, the optimal threshold will be determined by the tuning system. In tuning systems with larger intervals, this process may even be ignored completely, as the note transitions should be clearer.

2.2 *Sprechgesang* Analysis

The pitch segmentation performed on Hair's microtonal pieces described above are designed to find the note onset and offsets in music with discrete note-events in the presence of interpretive inflections. When the music has fewer discrete note-events, the fundamental assumptions underlying the iterative segmentation algorithm fail. In the extreme case of *sprechgesang*, it is less meaningful to discuss a segmentation in terms of "note onsets and offsets": the pitch trajectory becomes the important musical feature.

Our (ongoing) study into characterisation of pitch trajectory required a perceptually significant basis for comparison to be found in a pitch contour which varies continuously. A taxonomy of pitch segmentation[8] has previously divided the existing approaches into three: top-down; bottom-up and sliding-window. Most of the methods described rely on every point being checked to find where a candidate pitch segment can be divided. Once decided, the segment boundaries are never adjusted.

Our algorithm approaches the problem of segmentation using LVQ[18] which not only removes the need to check every point (instead, it iterates towards the best fit line by locally re-evaluating a linear regression) but also continually re-evaluates every split point, a property which is particularly useful if any meaningful, analytically useful segmentation of *sprechgesang* is to be achieved. Specifically, the algorithm deployed here uses k-means vector quantisation using a biased 2D Euclidean norm $||\mathbf{v}||_b = \sqrt{(\mathbf{v}\cdot\hat{\imath})^2 + b^2(\mathbf{v}\cdot\hat{\jmath})^2}$. Varying the bias value b permits the relative significance of pitch and time offsets to be adjusted insofar as they pertain to the segmentation process.

The k-means VQ algorithm used in this study starts by considering the whole pitch/time dataset as a single segment, then proceeds as follows:

1. Perform a biased linear regression on the each dataset, finding a lines of best fit (LoBF).

2. Bisect the LoBF with the highest (biased) error.

3. Re-partition the data into new datasets by associating them with the closest LoBF using the biased Euclidean norm as a distance metric.

4. Repeat until an acceptable error is achieved.

3 Results

3.1 Discrete-Pitch Analysis

To assess the performance of the segmentation algorithm, the output was tested against manual segmentations of the performance. Three listeners were asked to manually segment a performance using the Audacity audio editor [1]. The individuals were to identify segments which exhibited a steady pitch. The onset and offset of each segment were determined by repeated auditioning and adjustment of the start and endpoint. The listeners performed segmentation without the use of the score. All of the test subjects were competent musicians. However, a comparison of the results shows that even human listeners have difficulty in segmenting a performance in a microtonal rehearsal situation.

There were two situations where manual segmentations disagreed. In the first situation, a fluctuation in pitch during a transition (which did not correspond to notes in the score) was judged by one listener to contain two very short notes, whereas the other two listeners judged the fluctuation to be too unstable to correspond to notes. The algorithm identified one segment within this fluctuation.

At the second point of dispute, two listeners judged one note to be two distinct notes. The remaining listener and the algorithm both identified one segment at this point. Besides these cases, the algorithm identified three extra segments, and failed to identify one segment. This section corresponded to one single note within the score. There is a discernable change in pitch, although whether this change is sufficient to constitute two separate notes is obviously contentious. To allow comparison across all subjects, the segments which corresponded to disputed sections were removed from the calculations in every segmentation list.

The mean absolute deviation across all common onsets and endpoints identified by the human listeners was 0.016 seconds and the maximum deviation found was 0.173 seconds. The mean absolute deviation of the algorithm's estimated times from the average human estimated time was 0.076 seconds, with a maximum at 1.2 seconds. Excluding this abnormality, caused by an extra segment which was removed from calculations, the maximum deviation is 0.720 seconds, and the mean absolute deviation is 0.072 seconds.

This is not considered to be a conclusive evaluation of the performance of such an algorithm. The listeners reported difficulty in discriminating segments based solely on pitch. They found difficulty in distinguishing between deviations in pitch and deviations of other audible qualities such as timbre and dynamics The graphical representation of audio information at the resolutions required for such a task also includes visual cues which may accompany a change in dynamics, pitch or timbre. The process of isolating an audio segment may also have psychoacoustic effects caused by removing the surrounding context.

However, where manual segmentations agreed on the macro scale, the deviation across listeners was actually quite low. The difference between *how* the performance was segmented i.e. which notes occurred rather than their exact times, demonstrates the difficulty of the task. The definition of a 'correct' segmentation in this context is difficult to determine.

A conclusive study would involve a psychoacoustic investigation of exactly where the perceived onsets and endpoint occur in passages which exhibit both vibrato and legato, and would certainly contain considerably more performances and listeners. Such a detailed analysis is outwith the scope of this study.

However, the aim of this project was not to perform accurate transcription. Rather, the aim of this project was to create a tool which would provide accurate analysis of performed pitch. An analysis of only the onsets showed that the algorithm's estimated onsets exhibited a mean deviation of 0.02 seconds from the average human predicted onset. The same comparison regarding the endpoints show a mean deviation of -0.06 seconds. Thus, the algorithm has a tendency to shorten the boundaries of note candidates at both ends. This conservative estimation of the boundaries of a note is desirable for segmentation for the purpose of analysing pitch accuracy because the aim is to find the steady section of a note rather than the perceptual onset and endpoints.

During the early stages of rehearsal, the soprano Amanda Morison and clarinetist Ingrid Pearson had the CMT's microtonal rehearsal software made available to them. Recordings of the vocal part were used for this analysis. The results of the performance analysis were recorded in PML[6, 7], along with the score. A matching algorithm was developed[10] and used to link the notes in the performance to their corresponding score notes. Using this information, an annotated score can be generated which provides a simple visualisation of pitch accuracy. Figure 5 shows a score annotated with pitch error. Where a performance note correctly corresponds to a note in the score, the performance accuracy is denoted above the score. The bars represent pitch accuracy plotted in the range ± 31.579 cents (one 19^{th} of an octave), which is to say that one grey line represents a pitch error of approximately 6 cents.

Notes which do not have a corresponding note in the performance are displayed with red noteheads. As can be seen in the score, there are two sections where there are a significant number of wrong notes: bars 8 & 9 and the final three bars in the performance. These sections are caused by pitch drift, where the performer has gradually lost her tonal reference. This has been identified as one of the main problems of microtonal rehearsal, as both the intervals themselves have changed, but also, in the case of 19ET at least, only A remains at the same frequency. All other intervals other than the octave are different from the 12ET system. Therefore the performer has few familiar references.

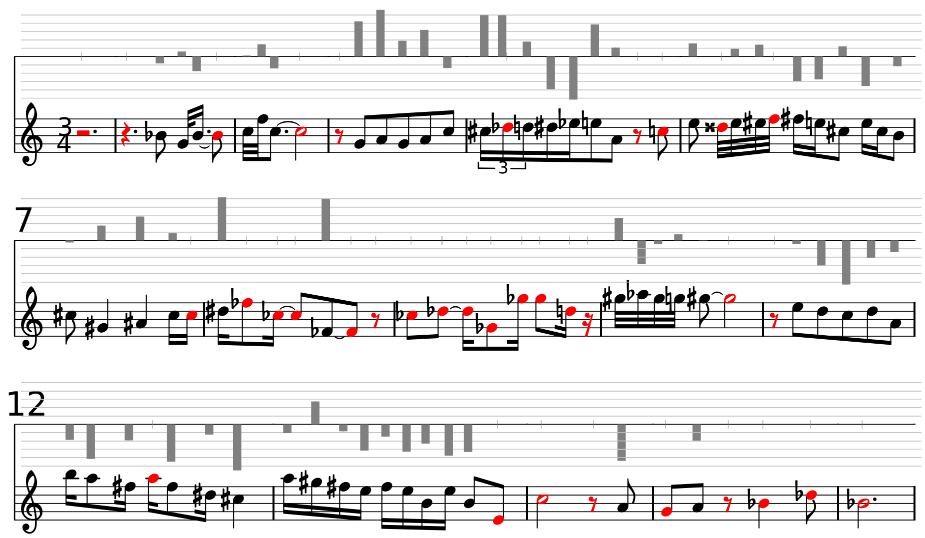

Figure 5: Annotated score displaying the deviation from exact frequency in a performance of the soprano part of Ash (composer Prof. Graham Hair).

4 *Sprechgesang* Analysis

As is the case in the microtonal study presented above, for the purposes of performance-musicological study, it is desirable to extract the expressive performance information contained in the pitch-transitions, and to separate regions of transition from regions of expressive note articulation. A first step towards these goals is to produce a nominal pitch-trajectory template; the deviations from the coarse-scale template, when referred via PML to the sheet music, will go far towards extracting significant articulation due to the performer's decisions from the less prescriptive score. To this end, it is desirable to produce a data-reduced characterisation of the pitch trajectories as automatically as possible.

Figure 6 shows the results of analysing the 2$^{\text{nd}}$ and 28$^{\text{th}}$ notes of the vocal recording made by Jane Manning as part of an on-going study into performance practice in Pierrot Lunaire. The pitch information was extracted from the vocal performance using the speech analysis software Praat[3]. This software also has the functionality to extract phonetic information from the performance, but the data used in this presentation is solely the fundamental frequency. The movement was manually segmented and analysed by Praat, and the VQ algorithm applied to produce a characterisation of the pitch trajectory. The algorithm was not constrained to segment three line-segments for each selection shown here; that three were chosen in both cases is coincidental.

The bias b for these experiments has been set at a value of 0.5. This is somewhat arbitrary and is based upon emperical observations that this does indeed produce accurate pitch trajectory characterisations throughout this movement. The fact that b has been chosen so arbitrarily is less than satisfactory, as is the inability of the sophisticated audio-score alignment algorithms presented in [10] to perform their task automatically. However, the charactisation of the pitch trajectories has been successfully achieved using a VQ method, and this is be-

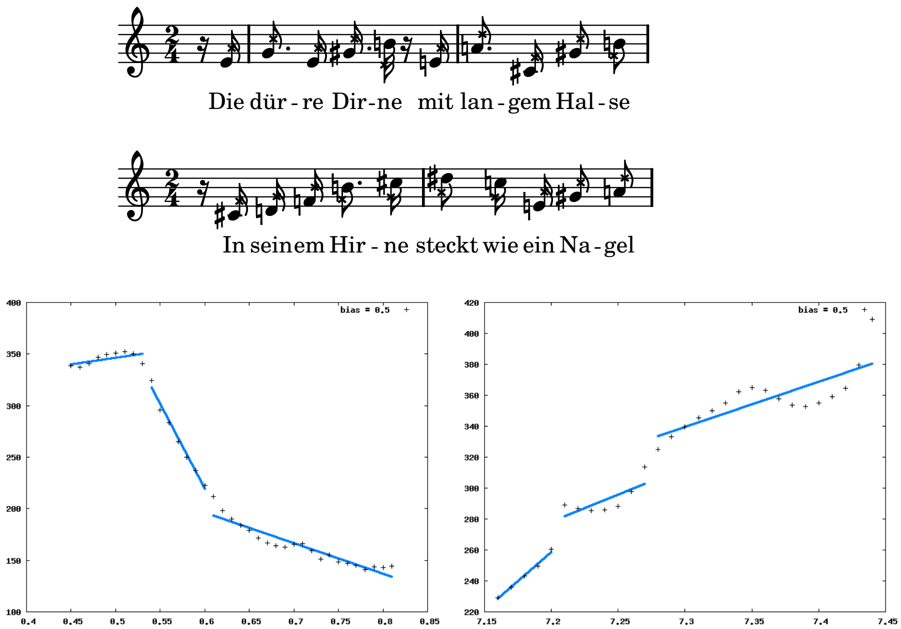

Figure 6: Notes 2 and 28 from a VQ-based analysis of Schoenberg's Pierrot Lunaire, No 12: "Galgenlied" ("Gallows Song") in context. Note 2 sets the first syllable of 'dürre'; note 28 the first of 'Nagel'.

ginning to provide useful information for our ongoing studies into performance practice in Pierro Lunaire.

5 Conclusions

The first pitch segmentation algorithm presented here, together with sequencers with enhanced microtonal functionality, have proved a useful rehearsal tool for performers of the highest levels of musical ability. The whole system seems to be accurate within the context in which it was tested, however it would benefit from further testing. The segmentation algorithm has not been directly tested against other segmentation algorithms. The segmentation algorithm presented by Rossignol *et al*[14] has no test data. The paper which succeeds that paper ([15]) gives results, but the test data is not specified, so a direct comparison is not possible. In any case, the aims of that segmentation algorithm is different from the one presented here. Rossignol *et al* were attempting to segment features such as phonemes in addition to changes in pitch, and the pitch trajectories in that paper do not seem representative of the microtonal rehearsal situation because the transitions are clear and notes seem to be obviously delineated by their pitch trajectory. Similarly, no results are presented for the transcription algorithm due to McNab[12] with which the current algorithm can be compared. However, the results of the algorithm here are consistent with segmentation performed by expert listeners.

Our new algorithm successfully segments a vocal performance for analysis of pitch accuracy in 19 tone equal temperament. The algorithm is not restricted to this tuning, however further testing may provide a deeper understanding of the parameters of the algorithms and improve performance in other tunings. The gradient threshold should be adjusted depending on the smallest chromatic interval present in the scale. For example, using 12 tone equal tempered music, the algorithm could be expected to perform even better because the larger chromatic interval causes a more pronounced pitch transition.

To improve usability, both algorithms would benefit from integration into a system which would provide the user with complete control over the analysis including the ability to select portions of the music and audition the performance at those points. This is possible immediately using the information within the PML file, and is awaiting the development of an adequate API for the representation for musical scores.

A system for the formal analysis of performance would require a rigorous study of the perceptual aspects, to ascertain exactly where onset and endpoints are perceived in complex situations involving gradual transitions involving pitch, vibrato and dynamics. Although this method is resistant to the presence of vibrato, the use of the vibrato 'removal' algorithm presented by Rossignol *et al*[13] may further improve performance and provide further analytical performance information.

The VQ pitch segmentation software we have developed is not yet as generally applicable as the note-based segmentation software, but in a problem-specific context is capable of producing useful results. Through the medium of PML, it may in future be possible further to automate the segmentation of *sprechgesang* by incorporating computer-derived phonetic information alongside marked-up score and audio resources.

References

[1] Audacity [computer program]. Home Page: http://audacity.sourceforge.net/ (accessed 28 October 2007), 2007.

[2] J. Bello, G. Monti, and M. Sandler. Techniques for automatic music transcription. In *International Symposium on Music Information Retrieval*, 2000.

[3] David Boersma, Paul & Weenink. Praat: doing phonetics by computer (version 4.6.12) [computer program]. Home Page: http://www.praat.org/ (accessed 28 October 2007).

[4] Mark Dolson. The phase vocoder: A tutorial. *Computer Music Journal*, 10(4):14–27, 1986.

[5] Graham Hair, Ingrid Pearson, Amanda Morrison, Nicholas Bailey, Douglas McGilvray, and Richard Parncutt. The rosegarden codicil: Rehearsing music in nineteen-tone equal temperament. *Scottish Music Review*, 1(1), 2007.

[6] A. R. Jensenius, A. Camurri, N. Castagn, E. Maestre, J. Malloch, D. McGilvray, D. Schwarz, and M. Wright. The need of formats for streaming and storing music-related movement and gesture data. In *Proceedings of the ICMC*, 2007.

[7] A. R. Jensenius, A. Camurri, N. Castagn, E. Maestre, J. Malloch, D. McGilvray, D. Schwarz, and M. Wright. A summary of formats for streaming and storing music-related movement and gesture data. In *Proceedings of the 4th International Conference on Enactive Interfaces (Accepted)*, 2007.

[8] Eamonn Keogh, Selina Chu, David Hart, and Michael Pazzani. An online algorithm for segmenting time series.

[9] Jane Manning. A sixties pierrot: A personal memoir. *Tempo*, 59(233):17–25, 2005.

[10] Douglas McGilvray. *On the Analysis of Musical Performance by Computer*. PhD thesis, The University of Glasgow, 2008.

[11] B. Mclaren. Microtonality in the u s, 2006.

[12] R. McNab, L. Smith, and I. Witten. Signal processing for melody transcription. In *Proc. 19th Australasian Computer Science Conf.*, pages 301–307, 1996.

[13] S. Rossignol, P. Depalle, J. Soumagne, X. Rodet, and J.-L. Collette. Vibrato: Detection, estimation, extraction, modification. In *Proceedings Digital Audio Effects Workshop*, 1999.

[14] S. Rossignol, X. Rodet, J. Soumagne, J.-L. Collette, and P. Depalle. Feature extraction and temporal segmentation of acoustic signals. In *Proceedings of the ICMC*, 1998.

[15] Rossignol S., Rodet X., Soumagne J., Collette J.-L., and Depalle P. Automatic characterisation of musical signals: Feature extraction and temporal segmentation. *Journal of New Music Research*, 28(4):281–295, 1999.

[16] M. R. Schroeder. Period histogram and product spectrum: New methods for fundamental-frequency measurement. *The Journal of the Acoustical Society of America*, 43(4):835–838, 1968.

[17] C. Weihs and U Ligges. Automatic transcription of singing performances. In *Bulletin of the International Statistical Institute, 54th Session*, pages 507–510, 2003.

[18] R. Gray Y. Linde, A. Buzo. An algorithm for vector quantizer design. In *IEEE Transactions on Communications*, pages 702–710, January 1980.