Poincaré Sections for Pitch Mark Determination

Martin Hagmüller and Gernot Kubin

Graz University of Technology, Austria,
hagmueller@tugraz.at, g.kubin@ieee.org,
http://www.spsc.tugraz.at

Abstract. An approach to pitch mark determination using Poincaré sections is presented. While speech has been interpreted in terms of nonlinear systems theory for quite some time, not much effort has been made to exploit this knowledge in the problem of pitch mark detection.
This algorithm uses nonlinear state space embedding and calculates the Poincaré section at a chosen point in state space. Pitch marks are then found at the crossing of the trajectories with the Poincaré plane in a certain neighborhood of the initial point. The procedure is performed frame-wise to account for the changing dynamics of the speech production system. First results show promising performance, comparable to the pitch marking algorithm used in Praat.

1 Introduction

For pitch–synchronous processing of speech, accurate pitch marks are essential. A particular challenge is the correct determination of pitch marks for dysphonic voices. On the other hand, having a reliable method for pitch marking available, this could be used for enhancement of rough pitch, by reducing the fluctuations of the fundamental period.

Accurate and robust methods for pitch detection are of interest for the analysis of dysphonic voices [1] and, e.g., for the measurement of jitter, methods to reliably determine the instantaneous fundamental period are necessary.

The nonlinear nature of the speech signal has been of increasing interest for several years now, starting in the early nineties [2]. Conventional algorithms, such as correlation based methods, assume linear models of speech production though, even for normal voices, these models cannot fully explain the properties of the signal. For dysphonic speech, these models more or less fail due to the higher order of non-linearity inherent in the system. Especially, for strongly irregular voices, conventional algorithms for pitch mark determination fail and, therefore, the need for new methods is at hand. Non-linear methods seem to be a promising way of overcoming the weaknesses of the currently used approaches.

State-space approaches for dysphonic voice analysis have been proposed recently [3, 4]. Voice irregularities have been treated with nonlinear methods before, e.g. by performing noise reduction in state space [5].

Depending on the application one wants to analyze either the periodicity or the harmonicity of the signal. This is specially of interest for irregularities

of the fundamental frequency, which can either be interpreted as a variation of the fundamental period in the time domain, or as an additional subharmonic component in the frequency domain. For time domain based analysis information would be lost, when only the fundament period related to the subharmonic would be measured. On the other hand, frequency domain based analysis, such as used for harmonic plus noise modelling, would loose information if the subharmonic would not be considered.

The paper is organized as follows. Section 2 will give some background and review existing algorithms for pitch determination in state space. In section 3 the proposed state-space approach for pitch marking will be introduced and the algorithm will be explained. Section 4 will show some results and finally section 5 will conclude the paper with a summary and an outlook.

2 Background and Related Work

A non-linear dynamical system can be embedded in a reconstructed state-space by the methods of delays [6]. According to Takens [7], the state space of a dynamical system can be topologically equivalently reconstructed from a single observed one-dimensional system variable. For a D-dimensional attactor it is sufficient to form a $M \geq 2D + 1$ state space vector. Later it was generalized by Sauer *et al.* [8] to $M > 2D_F$, where D_F is the (fractional) box counting dimension. The M-dimensional trajectory is formed from a speech signal vector $\mathbf{x}(n)$ by delayed versions of the signal $\mathbf{x}(n)$,

$$\mathbf{x}(n) = [x(n), x(n - \tau_d), ..., x(n - (M - 1)\tau_d)], \tag{1}$$

where τ_d is the delay time, which has to be chosen such as to optimally unfold the attractor. If one chooses an arbitrary point on the attractor in an M-dimensional space then one can create a hyper-plane which is orthogonal to the flow of the trajectories at the chosen point. This is called the Poincaré plane. All trajectories, that return to a certain neighborhood of the initial point, cross the hyperplane and can be represented in dimension $M - 1$ compared to the original trajectory.

In 1997 we [9] first suggested to use Poincaré sections for the determination of pitch marks and mentioned special applications for signals with irregular pitch period. Experiments showed very promising results for an example with vocal fry, where the pitch period doubles for some time. The pitch period was followed correctly.

Mann and McLaughlin [10] further worked with Poincaré maps and applied them to epoch marking for speech signals. They again saw promising results, but reported inability to resynchronize after, e.g., noise-like portions of speech as found e.g. in fricatives.

More recently, Terez [11] introduced another state space approach to pitch detection, using space-time separation histograms. The approach is very similar to the autocorrelation method, with sharper peaks, though. Since histograms are based on averaging statistics, localized pitch marks cannot be determined reliably with this approach.

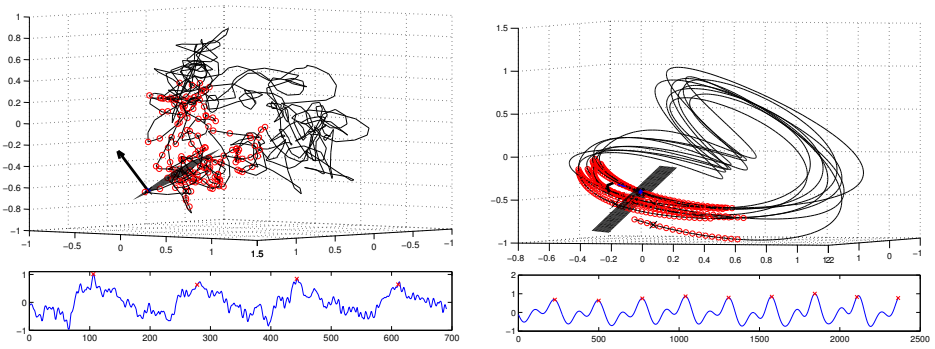

Fig. 1. *Top: Projection of state-space embedding on 3 dimensions and Poincaré plane. Circles are neighbors. Vowel 'a' from a dysphonic speaker. Bottom: Waveform Plot. Left: No low-pass filter. Right: Low-Pass Filter*

3 Description the Algorithm

This work builds on the before mentioned approaches and is an improved version of the work described in [12].

The algorithm works on a frame-by-frame basis to handle the slowly changing parameters of the speech production system. For pitch mark detection, the low-dimensional characteristics of the signal need to be observed. So the noise has to be removed, otherwise, specially for hoarse voices, the attractor is hardly visible with 3-dimensional embedding (Fig. 1 left). If the embedding dimension is high enough, intersections with the Poincaré plane would still be corrosponding to the pitch period, but with less reliability. At a certain noise level the algorithm breaks down, though. For a noise reduced attractor a singular-value-decomposition (SVD) embedding approach has been proposed [10], but similar results can be achieved by a simple linear-phase low-pass filter. The latter is computationally less demanding of course, so this is chosen for noise reduction. Note, the linear-phase property of the filter is necessary to preserve the time relationships of the signal.

Then the signal is upsampled to $f_s = 48$kHz to increase the resolution of the pitch marks, since at low sampling rates the pitch marks would exhibit too much discretisation error. The embedding in the state space is done by the method of delays, the embedding dimension was chosen to be $M = 8$. Experiments showed that this number gives the must robust results over different kind of speech sample, though the algorithm will work with embedding dimensions $M \geq 3$.

Poincaré section

At the heart of the algorithm is the calculation of the Poincaré hyperplane (Fig. 1 right). Around a chosen point $\mathbf{x}(n_0)$, the neighborhood $\mathcal{N}(n_0)$ is searched for the k closest points, according to the euclidian distance measure. Then a mean

flow direction $\mathbf{f}(n_0)$ of the trajectories in this neighborhood $\mathcal{N}(n_0)$ is calculated (considering only those trajectories, with a flow in the same direction as the initial point).

$$\mathbf{f}(n_0) = mean(\mathbf{x}[n+1] - \mathbf{x}[n]) \qquad \forall n \in \mathcal{N}(n_0), \tag{2}$$

where only trajectories are considered, which point roughly in the same direction as the initial flow vector ($\mathbf{f}^0[n]\mathbf{f}^0[n_0] > 0.9$, where $\mathbf{f}^0$ is a unit-length vector), i.e., orthogonal flow vectors or flow vectors in the opposite direction will not be considered.

So for every frame the Poincaré hyperplane is defined as the hyperplane through $\mathbf{x}[n_0]$, which is perpendicular to $\mathbf{f}[n_0]$ (Fig. 1 right).

The points, which are at the intersection of the trajectory with the Poincaré plane are considered as pitch marks. The exact location of these points is calulated by linear interpolation between the two points left and right of the intersection.

The length of one frame is chosen so that at least two periods of the expected minimum frequency fit into the frame. If the signal is quasi-periodic, the trajectory returns at least once into the chosen neighborhood and intersects the Poincaré hyperplane and a pitchmark can be detected. The hopsize depends on the the pitchmark in the current frame. The beginning of the following frame is set to the last pitchmark.

The voiced/unvoiced decision is another important issue. First, a frame is considered as unvoiced if the energy of the low-pass filtered signal is below a certain threshold. If the energy criterium does not detect an unvoiced frame it is further analyzed in the state-space. A first hint is the presence of high fluctuations in the fundamental periods in one frame. In this case the euclidian distance of the pitch-mark candidate points to the query point is considered. If this distance is over a certain threshold the points are discarded. Additionally, points, whose distance is much more than the mean distance of the candiate points to the initial point, are also discarded. This also reduces the occasional hit of a first harmonic.

If in the given frame of minimal and maximal expected fundamental period there is no intersection with the Poincaré plane, or if too many sections within the possible time frame are present, the signal is considered as stochastic and therefore unvoiced.

We have not been able to avoid an occasional phase drifting of the pitchmark. Specially for applications where high accuracy is an important issue this maybe a major disadvantage of this method.

Pseudo Code

```
· Low-pass filter
· Upsample (if necessary)
· WHILE index < lastindex - framelength
    · Get segment with framelength at index
    · IF energy(segment) < threshold,
```

```
        · set frame unvoiced
        · take next frame
    · END
    · Choose initial point (x(n_0))
    · Embed segment in pseudo state space (dimension, delay)
    · Select k neighbors in state space neighborhood N(n_0) of x(n_0)
    · Compute estimate of average vector flow f(n_0)
    · Define Poincaré hyperplane perpendicular to f(n_0) going through x(n_0)
    · Calculate intersection of trajectories through Poincaré plane by interpolation
      between samples neighboring Poincaré plane
    · IF std(T0) > 0.15 median(T0)
        · discard points far away from x(n_0)
    · END
    · Set index to beginning of next frame
· END WHILE
```

4 Results

Formal evaluation of the pitch marking problem still has to be performed. Informal results using the pitch detection evaluation database by Paul Bagshaw [13] (http://www.cstr.ed.ac.uk/projects/fda/) and disordered voice recordings are very promising.

In figure 2 the results of the algorithm on running speech can be seen. The German sentence *'Nie und nimmer nimmt er Vernuft an'* is spoken by a male speaker with modal voice. Most of the pitch marks are correctly set.

In figure 3 a segment is taken out of a speech file. There, a short period of diplophonic fundamental frequency is present (sentence *'rl040'* from the Bagshaw database). Other algorithms like Praat [14] fail at this instance or detect a period doubling if the chosen minimum pitch value allows for such long pitch periods. The Poincaré method recognizes the rapidly alternating pitch period correctly. Though in this case it is a matter of definition whether the alternating period or the period doubling is the correct interpretation, as was already discussed in the introduction.

Comparisons to the afore mentioned related algorithms (Section 2) are difficult, since none of the presented work deals with running speech, using full sentences. The analysis is restriced to voiced only sections of speech.

5 Conclusion

An algorithm using Poincaré sections for pitch mark determination for dysphonic voices was presented. The algorithm works in a pseudo state-space, using the method of delays for embedding. The results are very promising. Even at the current stage without any post processing the performance is comparable to the established pitch marking routine used by Praat, which uses sophisticated post processing. Specially the application for disordered voices will receive further evaluation.

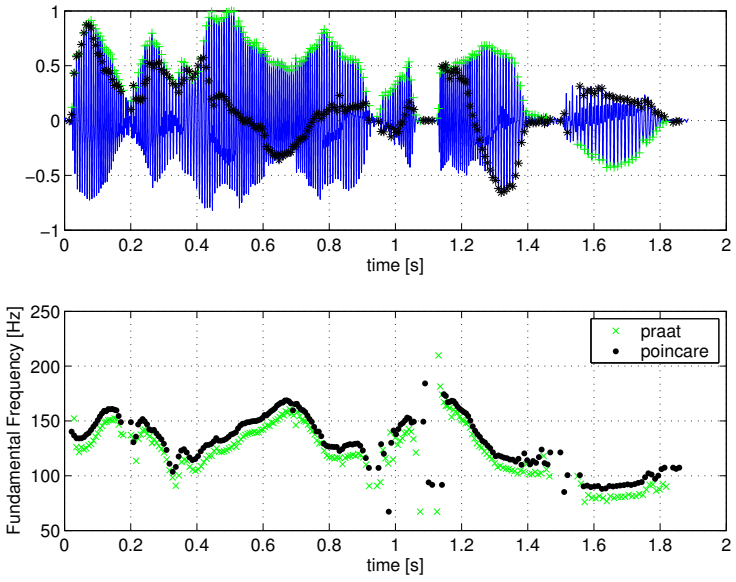

Fig. 2. *Top plot: Waveform plot of the sentence 'Nie und nimmer nimmt er Vernunft an' of a male modal voice and the pitch marks obtained by Poincaré section and Praat. Bottom plot: Fundamental frequency (The Praat result is biased by -10 Hz, for better visibility)*

References

1. Titze, I.R.: Workshop on acoustic voice analysis - summary statement. In: Proc. Workshop on Acoustic Voice Analysis, Denver, Colorado (1994)
2. Tishby, N.: A dynamical systems approach to speech processing. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. Volume 4., Albuquerque, NM (1990) 365–368
3. Matassini, L., Hegger, R., Kantz, H., Manfredi, C.: Analysis of vocal disorder in a feature space. Medical Engineering & Physics **22** (2000) 413–418
4. Herzel, H., Holzfuss, J., Kowalik, Z.J., Pompe, B., Reuter, R.: Detecting bifurcations in voice signals. In Kantz, H., Kurths, J., Mayr-Kress, G., eds.: Proc. Nonlinear Techniques in Physiological Time-Series-Analysis, Freital, '95, Springer Verlag (1996) 23–27
5. Matassini, L., Manfredi, C.: Noise reduction for vocal pathologies. Medical Engineering & Physics **24** (2002) 547–552
6. Kantz, H., Schreiber, T.: Nonlinear Time Series Analysis. Cambridge University Press (2003)
7. Takens, F. In: Detecting Strange Attractors in Turbulence. Volume 898 of Lecture Notes in Mathematics. Springer, New York (1981) 366–381
8. Sauer, T., Yorke, J.A., Casdagli, M.: Embedology. J. Stat. Phys. **65** (1991) 579–616
9. Kubin, G.: Poincaré section techniques for speech. In: Proc. of IEEE Workshop on Speech Coding for Telecommunication '97, Pocono Manor, PA (1997) 7–8

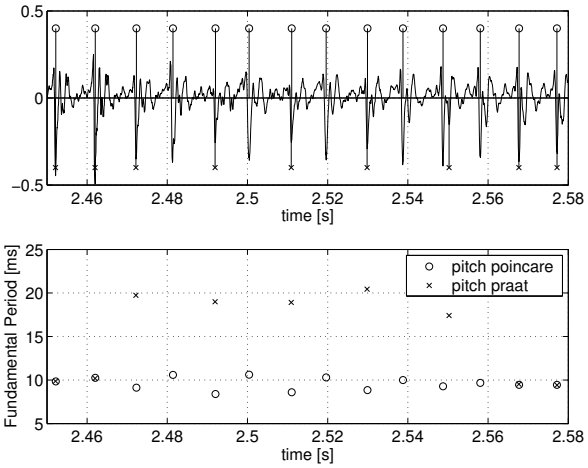

Fig. 3. *Top plot: Time-domain waveform plot with pitch marks with Poincaré section (positive peaks) and praat (negative peaks). Bottom plot: Fundamental period.*

10. Mann, I., McLaughlin, S.: A nonlinear algorithm for epoch marking in speech signals using poincare maps. In: Proceedings of the 9th European Signal Processing Conference. Volume 2., Rhodes Greece (1998) 701–704
11. Terez, D.: Robust pitch determination using nonlinear state-space embedding. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. Volume 1. (2002) 345–348
12. Hagmüller, M., Kubin, G.: Poincaré sections for pitch mark determination in dysphonic speech. In: 3rd International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA), Firenze, Italy (2003)
13. Bagshaw, P.: Automatic prosodic analysis for computer aided pronunciation teaching. PhD thesis, University of Edinburgh, UK (1994)
14. Boersma, P., Weenink, D.: (Praat, software, downloaded from http://www.praat.org, 10/2003)