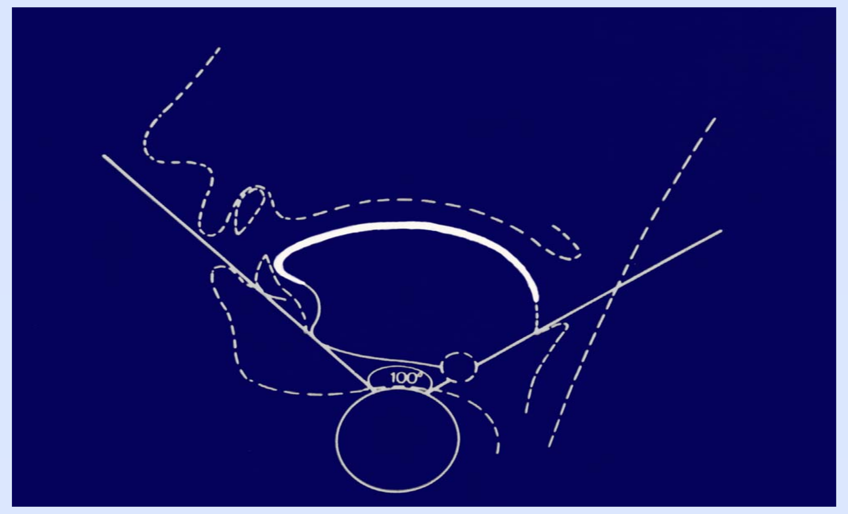

AN ALGORITHM FOR PSEUDO-3D REPRESENTATION OF THE CONTOUR OF THE TONGUE WHILE PLAYING THE DIDGERIDOO

Claudia Lindner, Wolfgang Angerstein, Volker Aurich, Thomas Massing

Inst. fuer Informatik, Univ. Duesseldorf, Universitaetsstr. 1 / Geb. 25.13, D-40225 Duesseldorf / Germany, cllindner@gmx.de
Phoniatric Dept., Duesseldorf Univ.-Clinic, Moorenstr. 5 / Geb. 13.77, D-40225 Duesseldorf / Germany, angerstein@med.uni-duesseldorf.de
Inst. fuer Informatik, Univ. Duesseldorf, Universitaetsstr. 1 / Geb. 25.13, D-40225 Duesseldorf / Germany, aurich-phon@acs.uni-duesseldorf.de
Acoustical Research Lab, Phoniatric Dept., Duesseldorf Univ.-Clinic, Moorenstr. 5 / Geb. 17.11, D-40225 Duesseldorf / Germany, massing@med.uni-duesseldorf.de

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

ABSTRACT

Introduction: In 1981 the importance of ultrasound images for the analysis of tongue movements was shown for the first time. As a result, algorithms have been developed for examining the ultrasound data with technological support. About a quarter century later the demand for a tongue contour segmentation algorithm based on current technological potential arose. Method: The developed algorithm processes mediosagittal B-mode images to a pseudo-3D representation of the tongue contour. Therefore, the processing steps smoothing, segmentation and pseudo-3D representation are applied to cross section image stacks. The main step (segmentation) deals with the detection, optimisation and completion of edges in grey value ultrasound images. In this context, pseudo-3D representation stands for the illustration of an image stack in the course of time rather than for a virtual three-dimensional representation. Results: Applying the algorithm to images recorded while playing the didgeridoo shows satisfying results for arbitrary ventral as well as dorsal views. But, under certain conditions artefacts might occur due to an unsatisfying signal-to-noise ratio. Conclusion: Results show improvement opportunities concerning technical optimisation like advanced transducer technology or further algorithm development. Moreover, the algorithm is applicable to other wind instruments and also for medical implications like sucking, swallowing or running speech.

INTRODUCTION

In 1981 Sonies et al. investigated the effectiveness of ultrasound images for the analysis of the movement of the tongue [Sonies, 1981]. The conclusion had been made that analysing the contour of the tongue is of high significance for the examination of tongue movements during running speech. At this time, the tongue contour was still drawn manually. Some years later, in 1988 Wein et al. presented a fully automatic computer-assisted solution for the segmentation of the contour of the tongue based on an ultrasound cross section image stack [Wein, 1988]. Moreover, the presented algorithm illustrated the extracted contour in a pseudo-3D representation. Since the algorithm was based on the technological standard back then (Atari), at present the request for a segmentation and representation algorithm based on the increased technological potential (PC) came up. For that reason, an algorithm has been developed for processing ultrasound B-mode images of the contour of the tongue with the intention of demonstrating the tongue movement in a pseudo-3D representation [Lindner, 2005]. In this context, pseudo-3D representation stands for the illustration of an image stack in the course of time rather than for a virtual three-dimensional representation. In terms of processing, ultrasound image sequences recorded while playing the didgeridoo are used.

METHOD

The algorithm is based on the processing of mediosagittal B-mode images of the tongue. These cross sections are recorded each 30 to 40 milliseconds. The procedure that generates a suitable pseudo-3D representation from a noisy ultrasound image sequence comprises three main steps: smoothing, segmentation and pseudo-3D representation. Firstly, the signal-to-noise ratio is improved by the application of appropriate smoothing algorithms. Afterwards, the contour of the tongue is segmented on the base of edge detection, and additional edge improvement algorithms are applied. Finally, the extracted contour is illustrated in a pseudo-3D representation.
The smoothing, as well as the segmentation, is done for each image separately, based only on the grey values of the image. Although the underlying algorithms are applied on single cross section images, the particular steps batch a complete cross section sequence. In the following an image *f* has always eight bit per pixel. That means the grey values range between 0, corresponding to the colour black, and 255, corresponding to the colour white.

Smoothing

Due to the underlying technique, ultrasound images can be very noisy as demonstrated in Figure 1. For that reason, the ultrasound images need to be pre-processed by applying appropriate smoothing algorithms. Since the segmentation is based on the smoothed images it is essential to preserve fine structures, in particular bright-dark edges. By the application of linear smoothing filters edges are blurred and get imprecise. Consequently, a well-defined border established between tongue and oral cavity would not be likely. Thus, edge-preserving smoothing filters like non-linear filters are needed. Aurich and Weule presented an edge preserving smoothing method which is based on non-linear modifications of Gaussian filters [Aurich, 1995]. For obtaining best possible results Aurich and Weule suggest using the demonstrated algorithms in a chain of three to five filters [Aurich, 1995; Weule, 1994].

Figure 1: B-mode cross section image of the contour of the tongue

With reference to the given ultrasound images the idea of using non-linear Gaussian filter chains has been adopted. Referring to this there are two different algorithmic approaches. The first one is named *Non-linear Separated Gaussian Filter* and the second one is named *Gaussian Weighted Median Filter*.
The *Non-linear Separated Gaussian Filter* is based on the non-linear Gaussian filter. It is defined as the composition of the following two filter steps.

$$NLGX_{\sigma,\eta} f(x,y) = \frac{1}{\alpha(x,y)} \sum_{u=-n}^{n} f(x+u,y) g_\sigma(u) g_\eta(f(x+u,y) - f(x,y)) \quad (1)$$

with $\alpha(x,y) = \sum_{u=-n}^{n} g_\sigma(u) g_\eta(f(x+u,y) - f(x,y))$

$$NLGY_{\sigma,\eta} f(x,y) = \frac{1}{\beta(x,y)} \sum_{v=-n}^{n} f(x,y+v) g_\sigma(v) g_\eta(f(x,y+v) - f(x,y)) \quad (2)$$

with $\beta(x,y) = \frac{1}{\beta(x,y)} \sum_{v=-n}^{n} g_\sigma(v) g_\eta(f(x,y+v) - f(x,y))$

σ and η are positive real parameters, *n* is determined by $n = floor(2\sigma + 1)$, $g_\sigma(t) = \exp(-t^2/2\sigma^2)$ is the unnormalised Gaussian function, $\alpha(x,y)$ and $\beta(x,y)$ and are factors that normalise the total weights to 1. Involving not only the location but also the input data *f* the non-linear Gaussian filter is edge-preserving and can even sharpen edges [Aurich, 1995; Weule, 1994]. The above *Non-linear Separated Gaussian Filter* yields quite similar results as the non-linear Gaussian filter published by [Aurich, 1995] and [Weule, 1994]. In terms of computing time it is much faster, and in our case it provides even better results for the subsequently applied segmentation algorithms.
The *Gaussian Weighted Median Filter* is based on the common median filter. Let $p = (x,y)$ be a pixel, and σ and *n* parameters chosen as above. Consider the neighbourhood $N = \{(u,v) : |u-x| \le n \text{ and } |v-y| \le n\}$ and assign $f(q)$ to each $q \in N$ with the weight w instead of the weight 1 as in the common median.

$$w = g_\sigma\left(\sqrt{(x-u)^2 + (y-v)^2}\right) \quad (3)$$

Sort the numbers $f(q)$ with $q \in N$ as ascending sequence $a_1, ..., a_{(2n+1)(2n+1)}$ with their corresponding weights $w_1, ..., w_{(2n+1)(2n+1)}$. Let W_p be the sum of all weights. Then the value $GM_\sigma f(p)$ of the *Gaussian Weighted Median Filter* with threshold $\gamma \in [0,1]$ is defined as the element such that:

$$\sum_{j=1}^{k} w_j \ge \gamma \cdot W_p \quad \text{and} \quad \sum_{j=1}^{k-1} w_j < \gamma \cdot W_p \quad (4)$$

In practice the *Gaussian Weighted Median Filter* is calculated line by line beginning in the top left corner, and already computed values are involved in successive calculations. Thereby the smoothing effect is enhanced. As already mentioned before, the non-linear Gaussian filter assures the best results by applying it in a chain. For that reason, the two described smoothing methods are applied alternately in a filter chain as follows:

$$(NLGY_{\sigma_3,\eta_3} \circ NLGX_{\sigma_3,\eta_3}) \circ GM_{\sigma_2} \circ (NLGY_{\sigma_1,\eta_1} \circ NLGX_{\sigma_1,\eta_1}) \quad (5)$$

The integration of the *Gaussian Weighted Median Filter* is not to be seen as a part of the chain itself rather than as a supportive intermediate step. However, the selection of the respective filter parameters influences the results, and provides scope for adaptation to several unique data sets. For a more detailed explanation of the filter parameters the reader is referred to [Lindner, 2005; Aurich, 1995; Weule, 1994].

Segmentation

The segmentation of the contour of the tongue is conducted in stages consisting of the detection, the optimisation and the completion of bright-dark edges. The three steps are built upon each other and are interdependent. Being the first segmentation step the edge detection algorithm processes the smoothed ultrasound images and lays the foundation for the following segmentation steps. The entire segmentation procedure is based on the algorithm *GetEdge* which is responsible for the detection of potential contour elements. The algorithm rests upon the assumption that the given images represent sagittal cross sections of the contour of the tongue. Due to the fact that the greater proportion of the tongue contour is horizontally oriented, the algorithm *GetEdge* postulates that the contour of the tongue is represented by distinctive horizontal bright edges. In this context, bright edge defines the transition from lower grey values to higher ones including the subsequent transition back to lower grey values. Moreover, it is assumed that the shape of the tongue contour is similar to that of a hill. Thus, no other significant bright edges are expected to be found above the contour of the tongue. Contrary, because of the musculature of the tongue, several significant bright edges can be found below the tongue contour (see Figure 2). According to the assumptions mentioned above the algorithm *GetEdge* passes through each image column by column in a top down manner and searches in each column for significant decreases of the grey values. In this regard the significance is determined by a parameter that sets the maximum allowed difference between the first brightest pixel in a column and successive lower grey values. When selecting this parameter it must be taken into account that a too low threshold can cause the misinterpretation of noise as a bright-dark edge, whereas a too high threshold can lead to missing out the tongue contour. The first brightest pixel in the x^{th} column that is followed by a significant grey value decrease is marked as an element of the contour of the tongue. Let *C* be the set of all contour pixels. This segmentation principle is visualised in Figure 2.

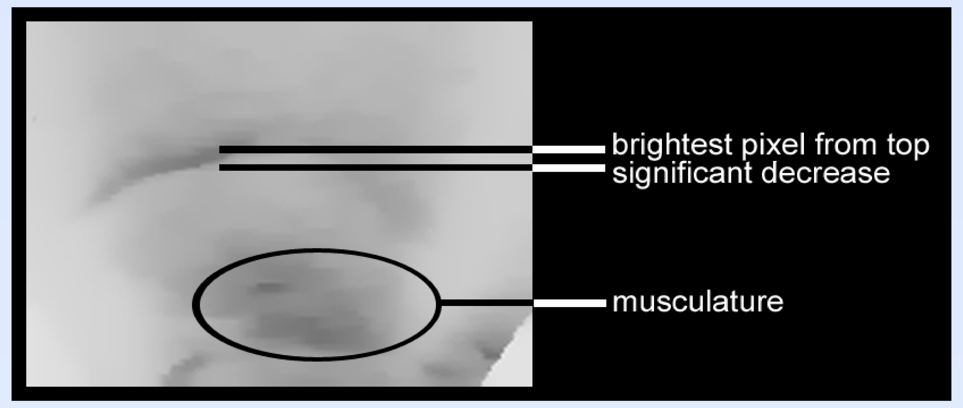

Figure 2: Significant grey value decrease

In the majority of the cases the edge detection alone does not provide satisfying results in terms of getting the best possible precise illustration of the contour of the tongue. For that reason, several optimisation algorithms have been developed with the intention of trying to identify pixels that have been classified as contour pixels within the scope of edge detection but which are not. There are two basic optimisation algorithms as well as contour specific optimisation algorithms. Latter can be added in a user-defined composition to the particular chosen basic optimisation algorithm. The basic optimisation algorithm *OptimizeEdgeNeighbxy* investigates whether every identified contour pixel belongs to an uninterrupted part of the contour. Pixels that are located aside from the actual tongue contour are regarded as a misinterpretation, and consequently eliminated from set C. For updating the set of contour pixels choose a rectangle *N* centred at (0,0), and eliminate all pixels p from C for which p+N contains at least one column without pixels in C. The alternative basic optimisation algorithm *OptimizeEdgeMinLength* is the extended version of *OptimizeEdgeNeighbxy*. Whereas latter considers the complete vertical as well as horizontal neighbourhood for searching for continuation pixels, *OptimizeEdgeMinLength* investigates the potential contour pixels only in relation to their horizontal continuation qualification. Pixels that cannot be continued up to a minimal length of $L_{\min} > 0$ are regarded as a misinterpretation. Thus, eliminate $p = (x,y)$ from C, except if there exist pixels $(x+1, y_1), ..., (x+L_{\min}-1, y_{L_{\min}-1}) \in C$ such that $|y_j - y_{j-1}| \le n$ for $j = 1, ..., L_{\min}-1$ and $n > 0$.
In the context of tongue contour specific optimisation the acquired potential contour pixels are analysed in relation to significant features. In this case, significant features mean pixels that fulfill certain requirements in relation to the fact that the given images display the contour of the tongue. Thus, contour specific knowledge is involved. The algorithm *OptimizeEdgePeak* includes knowledge about the hill shape of the tongue contour. Pixels, which are located above the top of the tongue respectively hill are eliminated (see Figure 3). Since the algorithm *OptimizeEdgeVale* takes also the hill-related shape into account it completes the results of *OptimizeEdgePeak*. But, in contrast it eliminates pixels that are located below the left and right marginal maximum. The third contour specific optimisation algorithm is called *GetPeakEdge*. It investigates the gradients on both sides of the top of the tongue. Pixels that excess the acceptable gradient difference are eliminated.

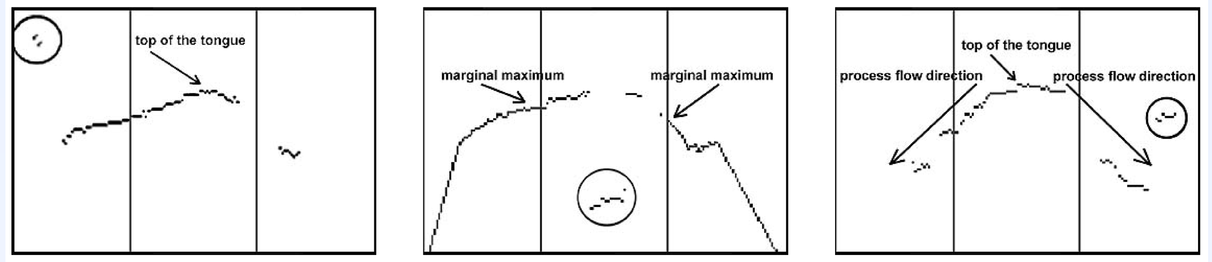

Figure 3: Processing scheme of the three contour specific optimisation algorithms

Depending on the quality of the images and the particular parameter selection there are in all likelihood gaps between identified contour pixels. These gaps need to be filled in order to get a complete illustration of the contour of the tongue. Therefore, an edge completion algorithm is used to bridge existing gaps. Being very efficient and sufficient for that purpose Bresenham's line algorithm has been chosen [Bresenham, 1965; Lindner, 2005].

Pseudo-3D representation

When evaluating the movements of the tongue with respect to several medical application areas it is essential to observe successive cross sections in their entirety. For that reason, the pseudo-3D representation illustrates whole tongue movement sequences by projecting successive temporary cross sections in the course of time. Since spatial effects are not that important for the pseudo-3D representation the parallel projection has been chosen as projection method with the time being the third dimension. Figure 4 demonstrates the exemplary results after applying the respective parallel projection.

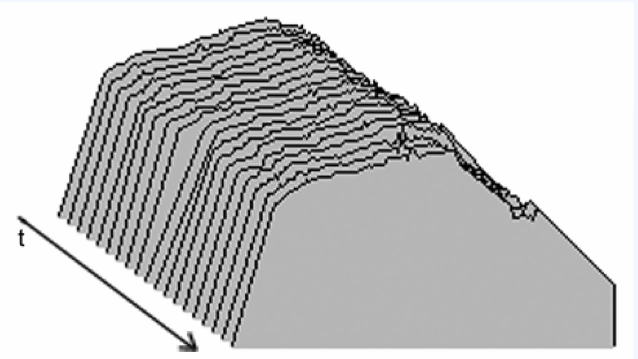

Figure 4: Pseudo-3D representation

RESULTS

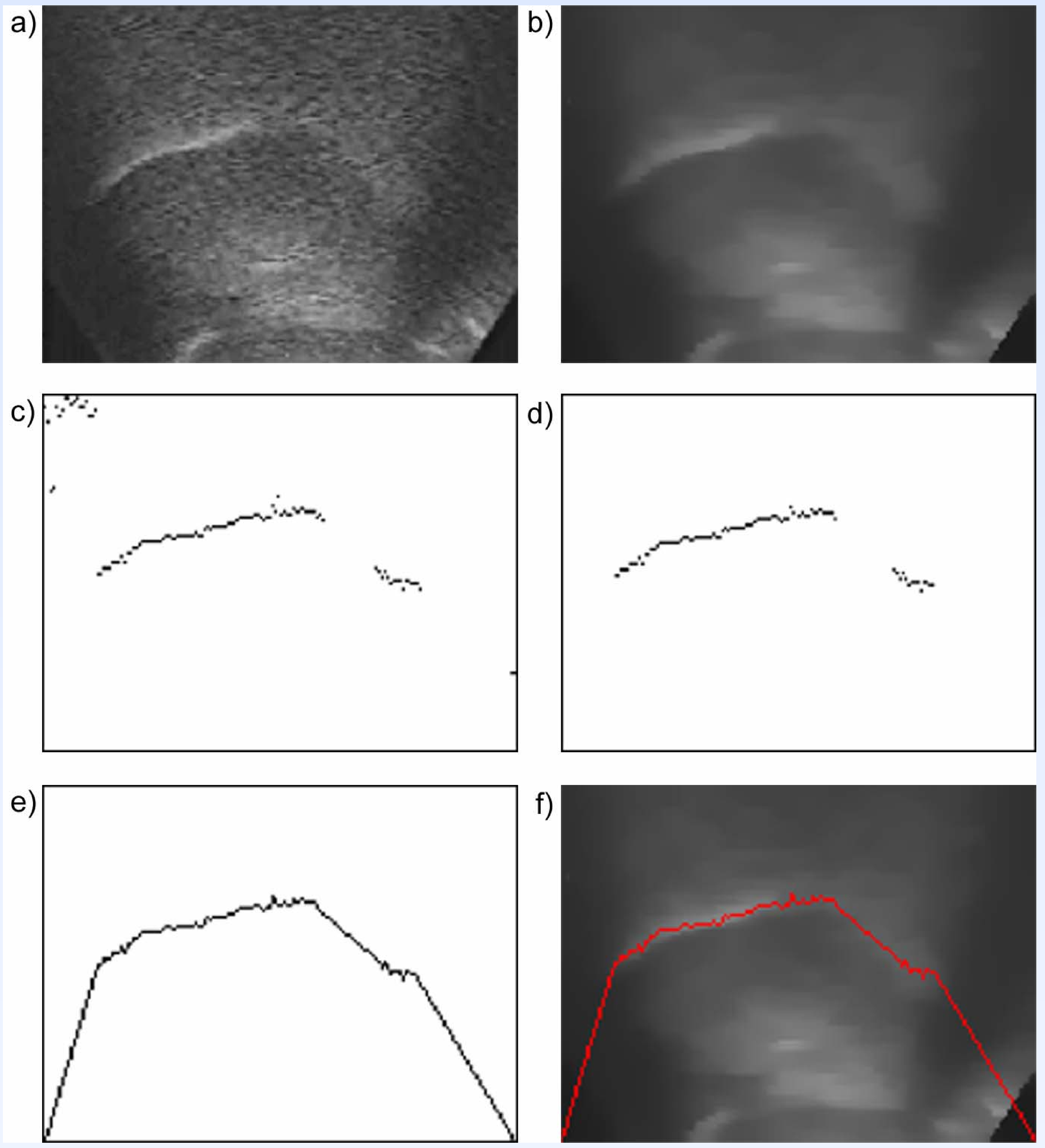

Figure 5: Step-by-step processing illustration:
a) original image, b) result after smoothing, c) result after edge detection,
d) result after edge optimisation, e) result after edge completion, f) final result in contrast to the original image

The described algorithms provide a process chain for obtaining a pseudo-3D representation of the contour of the tongue as illustrated in Figure 4. For achieving the requested results several algorithm steps are required. The effectiveness and efficiency of the particular algorithms can be controlled by optional parameters [Lindner, 2005]. Figure 5 gives a review of the above described algorithms considering the particular result after each step.
The effectiveness of the described algorithms was reviewed on the basis of ultrasound images recorded while playing the didgeridoo. The corresponding results are satisfying for arbitrary ventral as well as dorsal views. However, under certain conditions like insufficient dark-bright edges in the original ultrasound images, artefacts might occur on account of an unsatisfying signal-to-noise ratio.

CONCLUSIONS

Although the algorithms show satisfying results, there is still space for improvement opportunities in terms of technical optimisation. First of all, the signal-to-noise ratio can be enhanced by the usage of advanced transducer technologies. By providing ultrasound images of a higher resolution, this improvement approach will give a wide scope for the usage of the above presented algorithms. Moreover, further algorithm development could be done for improving the algorithmic effectiveness when processing images of a poor signal-to-noise ratio. Furthermore, at this stage the processing chain is not real-time capable. The reason is not only the batch processing, but also the particular algorithm related processing time for each image. Currently, the smoothing step is the most time-consuming part of the processing chain. But this time problem might be solved by using better quality ultrasound images. Finally, the visualisation of the pseudo-3D representation can be enhanced by the usage of numerical interpolation instead of Bresenham's line algorithm, and by displaying the tongue contours in colours along the colour spectrum.
In the context of medical application areas, the algorithm can be tested with other wind instruments as well as with ultrasound images recorded while sucking or swallowing, or during running speech.

REFERENCES

Aurich, V. ; Weule, J. (**1995**). "Non-Linear Gaussian Filters Performing Edge Preserving Diffusion", *17. DAGM-Symposium (Bielefeld)*, pp. 538-545
Bresenham, J.E. (**1965**). "Algorithm for computer control of a digital plotter", *IBM Systems Journal* 4(1), pp. 25-30
Lindner, C. (**2005**). "Segmentierung der Zungenkontur in Ultraschallaufnahmen", www.lbusz.de/scientificpapers/ClaudiaLindner_BScThesis.pdf
Sonies, B.C. ; Shawker, T.H. ; Hall, T.E. ; Gerber, L.H. ; Leighton, S.B. (**1981**). "Ultrasonic visualization of tongue motion during speech", *Journal of the Acoustical Society of America* 70(3), pp. 683-686
Wein, B. ; Alzen, G. ; Tolxdorff, T. ; Böckler, R. ; Klajman, S. ; Huber, W. (**1988**). "Computersonographische Darstellung der Zungenmotilität mittels Pseudo-3D-Rekonstruktion", *Ultraschall in der Medizin* 9, pp. 95-7
Weule, J. (**1994**). "Iteration nichtlinearer Gauß-Filter in der Bildverarbeitung", www.7b5.de/Weule_Thesis.pdf