Level-correct Speech Recordings for the Analysis of Parkinson Speech

Lena M. Jaschke[1], Christin Baasch[1], Gerhard Schmidt[1], Adelheid Nebel[2] and Günther Deuschl[2]
[1] *Dept. of Digital Signal Processing and System Theory, Kiel University, E-mail: {leja, chrb, gus}@tf.uni-kiel.de*
[2] *Dept. of Neurology, Kiel University, E-mail: {a.nebel, g.deuschl}@neurologie.uni-kiel.de*

Abstract

7 to 10 million is (according to the Parkinson's Disease Foundation [1]) the estimated number of people that suffer from Parkinson's disease worldwide. Thus, this disease is one of the most frequent neurodegenerative diseases in the world. Beside motor disorders, nearly 90 % of the patients suffer from a speech disorder, called dysarthria. Thereby the main problem is that the patients speak too quietly caused by a perception disorder, although they do not realize it. A variety of different voice therapies are the subject of current research. In order to be able to make reliable statements about the success of a treatment, one of the main aims is to make the voice recordings comparable. In this article, we focus on the implementation of a recording environment that enables voice recordings of patients with Parkinson's disease and measures the distance between the speaker and the microphones to apply a distance-based gain correction. This leads to distance-independent recordings. A subsequent analysis of these recordings should give a possibility for an evaluation of the effectiveness of different therapies on voice quality with respect to speech power. The main focus is on two localization algorithms and their performance in a real-time application to measure the distance between the sound source and a microphone array containing four sensors.

Introduction

By virtue of the huge number of people suffering from Parkinson's disease all around the globe and the speech disorders that tend to accompany this disease [2], the examination of the effect of different therapies on the speech of the patients is becoming a significant area of interest. While current evaluations in the German speaking setting are mostly based on auditive assessments and are followed by a huge time investment, the aim is to create a recording tool that evaluates the success of a therapy on the speech quality of a patient automatically and objectively.

As is well known, the level of audio signals depends on the distance between the audio source and the microphone. Therefore, in order to be able to compare voice recordings that have been made in advance of a therapy with those taken after a speech therapy, a measurement of the distance is necessary. An adjustment of the audio signal afterwards based on the calculated distance and a reference distance value will make it possible then to make reliable statements about the success of a treatment.

Two different algorithms were implemented, with the goal to calculate the distance between the acoustic source and a microphone array for distances between 20 cm and 60 cm as precisely as possible and with a small number of microphones.

The recording environment (shown in Fig. 1) contains an USB audio interface with four gain-controllable input channels as well as a microphone array with four condenser microphones placed in a square and a processing environment. The implementation of the algorithms takes place in a general-purpose real-time processing toolkit written in C/C++.

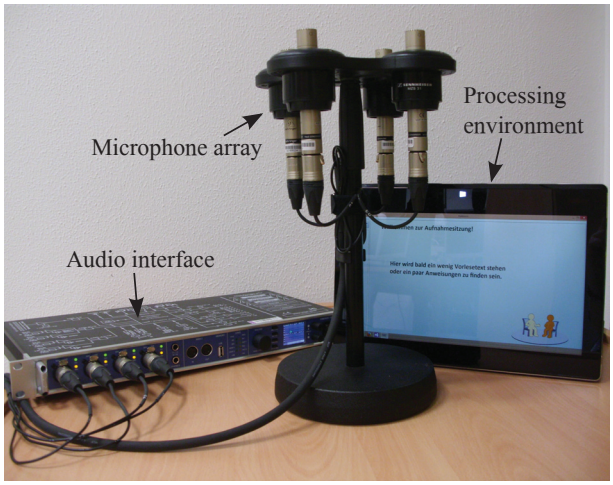

Figure 1: The recording environment containing a microphone array with four condenser microphones, the USB audio interface and a processing environment.

Our research investigates an efficient way to measure the distance between a source of sound and a microphone array combining two different acoustic source localization algorithms. All the calculations will be achieved in the time domain without the use of spectral information.

In the following, we will briefly present the general groups of localization algorithms, then we present the two implemented localization algorithms and compare them.

Sound Localization Methods

In this article, the term localization refers only to the estimation of the distance between a source of sound and a microphone, and not the actual position of the acoustic source. There are two superordinate groups of localization algorithms — Time Difference of Arrival (TDoA) and Received Signal Strength (RSS) algorithms. Most existing techniques use the information from various microphones for their calculation. Any number of micro-

phones operating in tandem independently of their exact arrangement are known as a microphone array.

TDoA-based localization algorithms

TDoA localization algorithms — according to their name — use time differences of the signals arrival at each microphone for the calculation. In order to calculate these time differences the most straightforward method is to determine them using the cross correlation between the two signals. The cross correlation function (CCF) $\hat{r}_{y_i y_j}(\kappa, n)$ of the two microphone signals $y_i(n)$ and $y_j(n)$ can be achieved by summing the signals for every possible delay between the two microphones. The estimated time delay then will be represented by the lag where the maximum of the cross correlation function can be found. The recursive calculation of the cross correlation function in the time domain is defined as follows

$$\hat{r}_{y_i y_j}(\kappa, n) = \beta \cdot \hat{r}_{y_i y_j}(\kappa, n-1) + (1-\beta) \cdot y_i(n) \cdot y_j(n+\kappa). \quad (1)$$

As displayed in Eq. (1) — for efficiency reasons — it is not necessary to compute the whole CCF. Rather a calculation of the resulting lags for delays between the two considered signals, in the range of ± the maximum possible retardation, is sufficient. The maximum lag depends on the existing array arrangement and dimensions. Reflections and multipath propagation will not be the object of further consideration.

RSS-based localization algorithms

The second group — known as RSS localization algorithms — is based on the received signal strengths. As is well known, the strengths of an audio signal $p_i(n)$ decreases with its distance to the source of sound $d_i(n)$, as in Eq. (2) shown

$$p_i(n) \sim \frac{1}{d_i(n)}. \quad (2)$$

It follows the relation, shown in Eq. (3), between the distances $d_i(n)$ and $d_j(n)$ of two microphones to the source of sound and their associated levels $p_i(n)$ and $p_j(n)$

$$\frac{p_i(n)}{p_j(n)} = \frac{d_j(n)}{d_i(n)}. \quad (3)$$

Distance estimation for normalized voice recordings

The main focus of our work was to atain distance normalized voice recordings, as schematically illustrated in Fig. 2. Concerning this task the implemention of various subcomponents was necessary. Within this paragraph we will give a brief description of the preprocessing for the speech signals and we present the two localization algorithms implemented in a real-time framework.

Signal preprocessing

In anticipation of calculating the distance by the use of a localization algorithm, the influence of noise effects on the results get restricted by appropriate preprocessing of the speech signals.

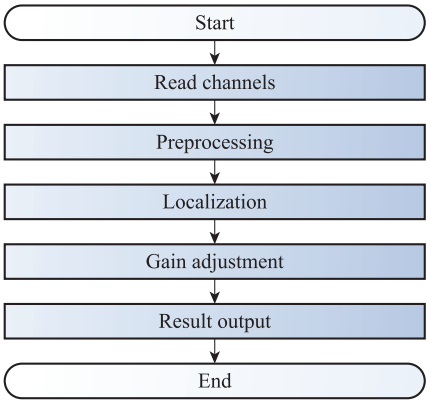

Figure 2: Schematic of the individual steps of the algorithm.

First the recorded signals are filtered using a band-pass filter with a lower cutoff frequency f_L of 200 Hz and an upper cutoff f_H of 10 kHz. The block size that is continously fetched from the hardware is 256 samples at a sample rate of $f_s = 44.1$ kHz.

In addition to filtering the signals, a simple algorithm is used to detect voice activity. To distinguish speech pauses from periods where speech activity takes place, the estimated background noise is compared with the current sound level. Once voice activity is detected, the block of filtered data is fed to the localization algorithms that are described in the following section.

Acoustic localization

The main purpose of the localization algorithms is to calculate the distance between the source of sound and the microphone array. Therefore two algorithms were implemented, whose calculations are based on different informations.

Level-based algorithm

The level-based algorithm represents a combination of an RSS and a TDoA algorithm. It uses the time delay between two microphones, as well as their received signal strengths, to calculate the distance value. A general overview of the individual components and used variables is given in Fig. 3. Beside the two microphones M_0 and M_1, which deliver the algorithm with the required informations to compute, the source of sound Q, as well as the distances $d_0(n)$ and $d_1(n)$ and the path difference $d_{\Delta,1}(n)$ are shown.

The functionality of this algorithm is restricted to cases of nonequal distances between the two considered microphones and the source of sound. In advance — in order to avoid this scenario — a channel selection takes place that is based on the maximum detectable lag between all combinations of the microphones contained by the array.

Once two appropriate channels are selected, the time difference $t_{\Delta,1}(n)$ of the signals arrival at each of this microphones can be used to calculate the present path difference $d_{\Delta,1}(n)$ in meter as in Eq. (4) displayed. The time difference $t_{\Delta,1}(n)$ again can be obtained due to the use of Eq. (1). The parameter c_0 represents the speed of sound in air, assumed to be 343,2 m/s.

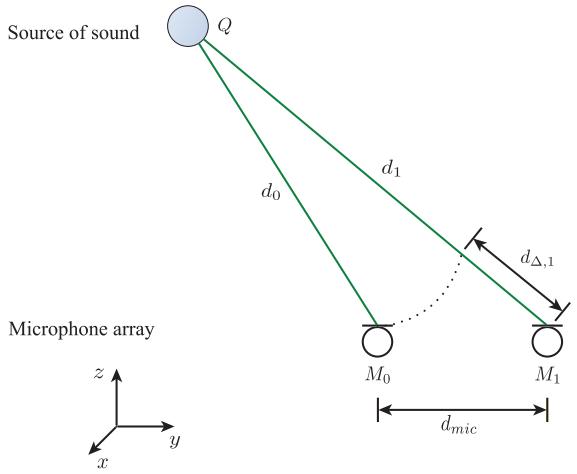

Figure 3: The components used in the Level-based algorithm. The index n is omitted for better reading visibility.

$$d_{\Delta,1}(n) = t_{\Delta,1}(n) \cdot c_0 \tag{4}$$

Furthermore $d_{\Delta,1}(n)$ can be expressed as the difference between the two distances $d_1(n)$ and $d_0(n)$

$$d_{\Delta,1}(n) = d_1(n) - d_0(n). \tag{5}$$

Due to the high noise susceptibility of the microphone levels, the values get filtered by an first-order Infinite Impulse Response (IIR) filter before inserting them in the further steps of the algorithm.

The relation between two distances and the corresponding microphone levels introduced in Eq. (3) leads us to

$$d_0(n) = d_1(n) \cdot \frac{p_1(n)}{p_0(n)}. \tag{6}$$

After inserting Eq. (5) in Eq. (6) the following equation can be obtained

$$d_0(n) = d_{\Delta,1}(n) \cdot \frac{p_1(n)}{p_0(n) - p_1(n)}. \tag{7}$$

The searched distance $d_0(n)$ between the source of sound and the microphone M_0 thus depends only on the two microphone levels $p_0(n)$ and $p_1(n)$ and the present path difference $d_{\Delta,1}(n)$ between them. However it is unaffected by the distance between the two microphones d_{mic}.

GPS-based algorithm

The GPS-based algorithm belongs to the class of the TDoA algorithms as its calculation is based exclusively on time differences of the signals' arrival. While the previous introduced level-based algorithm only requires an array consisting of 2 microphones this algorithm utilizes the information of four microphones.

Fig. 4 serves to illustrate the different components and the variables used within this algorithm. Again the microphones are marked through the parameters M_0, M_1, M_2 and M_3, the source of sound via the parameter Q, the distances between the source and the single microphones with $d_0(n)$, $d_1(n)$, $d_2(n)$ and $d_3(n)$ and the existing path differences are marked with the parameters $d_{\Delta,1}(n)$, $d_{\Delta,2}(n)$ and $d_{\Delta,3}(n)$.

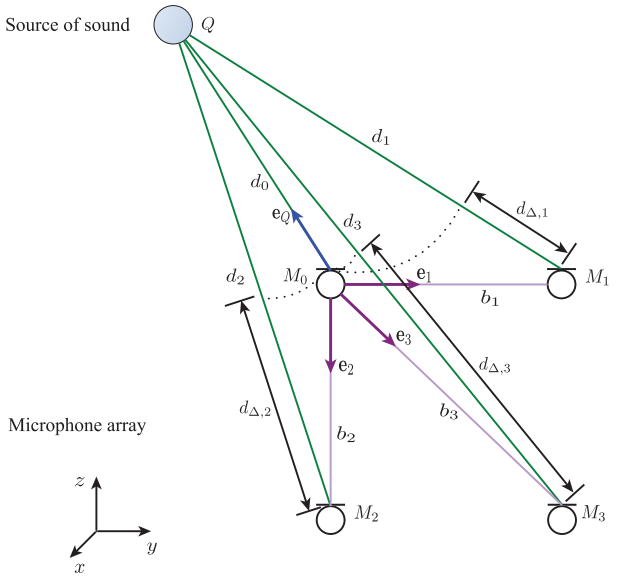

Figure 4: The components used in the GPS-based algorithm. The index n is omitted for better reading visibility.

We adopted this method from the "Analytical GPS Navigation Solution" introduced by Alfred Kleusberg in the year 1999 [3]. Therefore, we will only give a brief description of the single steps of this algorithm, for any deeper information we would refer at this point to the publication listed in the literature section.

In a first step — analogously as described within the previous section — we determine the path differences $d_{\Delta,i}(n)$ with $i \in \{1,2,3\}$ between the microphones M_1, M_2 and M_3 and the reference microphone M_0 as the product of the existing time difference $t_{\Delta,i}(n)$ and the speed of sound c_0

$$d_{\Delta,i}(n) = t_{\Delta,i}(n) \cdot c_0, \quad i \in \{1,2,3\}. \tag{8}$$

Based on the known cartesian coordinates of the microphones, the calculation of the spatial distances b_i with $i \in \{1,2,3\}$ between each microphone and the reference microphone M_0 as well as the calculation of the associated unit vectors e_i with $i \in \{1,2,3\}$ can be realised. With these quantities the distance $d_{0,1/2}(n)$ can be determined for every $i \in \{1,2,3\}$ according to

$$d_{0,1/2}(n) = \frac{1}{2} \frac{b_i^2 - d_{\Delta,i}^2(n)}{d_{\Delta,i}(n) + b_i[\mathbf{e}_{Q,1/2}(n) \cdot \mathbf{e}_i]}. \tag{9}$$

The variable $\mathbf{e}_{Q,1/2}(n)$ describes the unit vector pointing from the reference microphone M_0 in the direction of the source of sound Q. It should be noted that the number of feasible solutions provided by this method, depends on the array arrangement and the position of the source.

Gain adjustment

Once the distance $d_0(n)$ is calculated, either by the level- or the GPS-based algorithm, a gain adjustment takes place to achieve the goal of distance normalized voice recordings.

The previously calculated and by an first-order IIR filter smoothed distance values $\overline{d_0(n)}$ and a reference distance

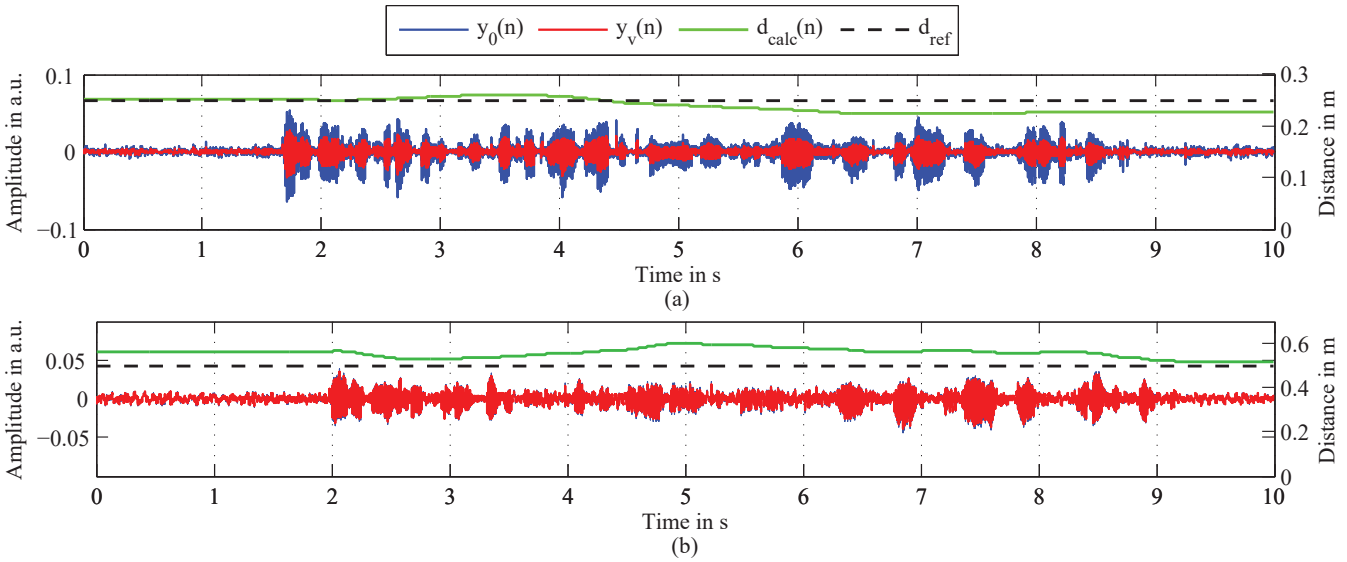

Figure 5: Calculated distance values $d_{calc}(n)$ for a combined use of the two algorithms with the weighting factors $W_{Level} = 0.4$ and $W_{GPS} = 0.6$. The output is shown for a reference distance value d_{ref} of 25 cm (a) and 50 cm (b).

value d_{ref} constitute the basis to determine the gain correction factor

$$V(n) = \frac{\overline{d_0(n)}}{d_{ref}}. \tag{10}$$

A distance normalized output signal $y_V(n)$ can then be obtained by a gain adjustment of the microphone input signal $y_0(n)$ with the smoothed gain correction factor $\overline{V(n)}$

$$y_V(n) = \overline{V(n)} \cdot y_0(n). \tag{11}$$

Experimental Results

Several tests were performed to compare the results obtained by this two approaches. We run through a number of experiments either with recorded sound samples or in real time to determine the performance of the methods under different circumstances. Beside varying the distance between the speaker and the microphone array, we examined the output for different array arrangements, smoothing parameters and noise conditions.

It has become apparent that the best outcome can be achieved by combining the two above described methods. Thus two weighting factors W_{Level} and W_{GPS} in the range between 0 and 1 were introduced. By summing up the product of each distance value calculated within the algorithm with its weighting factor, the result $d_{calc}(n)$ of this combined technique can be obtained.

Fig. 5 illustrates the calculated distances $d_{calc}(n)$ for a sound sequence with a duration of 10 s and the weighting factors $W_{Level} = 0.4$ and $W_{GPS} = 0.6$. For the upper case Fig. 5 (a) a speech recording with a distance d_{ref} of 25 cm between the speaker and the closest microphone forms the basis and the lower case Fig. 5 (b) is based on a distance d_{ref} of 50 cm. Furthermore the microphone input signal $y_0(n)$ and the gain normalized signal $y_V(n)$ are displayed. Within this experiment the background was composed of no added noise beside the present room noise, due to cooling fans.

A comparison of the calculated distance values with the reference value for each of the cases enables us to determine the accuracy of our method. Within ten seconds we observe a maximal deviation of 2.5 cm for a reference distance of 25 cm and 9 cm for $d_{ref} = 50$ cm.

Conclusion and Outlook

In this contribution, two different sound localization methods were proposed, in order to determine the distance between a source of sound and a microphone array and obtain distance normalized voice recordings. Experimental results show that the accurancy of the results is strongly influenced by noise effects. By introducing some restrictions and smoothing parameters, nontheless good results for the required application can be achieved. By combining the two algorithms we could raise the accuracy of the obtained results even more.

As we could achieve accurate results with a microphone array containing a low number of microphones — the presented level-based algorithm only requires 2 microphones — and the use of algorithms that demand a low grade of complexity, a large number of applications are conceivable. In our further research the recording tool will be used to obtain level-correct speech recordings of patients with Parkinson's disease.

References

[1] Parkinson's Disease Foundation: Statistics on Parkinson's, URL: `http://www.pdf.org/en/parkinson_statistics`

[2] Nebel, A. and Deuschl, G.: Dysarthrie und Dysphagie bei Morbus Parkinson. Georg Thieme Verlag KG, 2008

[3] Kleusberg, A.: Analytical GPS Navigation Solution. Technical report, University of Stuttgart, 1999