Observations from a Simple Vocal-Tract-Model's Behaviour for PD-Dysarthric Speech: Evaluations

Ulrich Heute

Dig. Sig. Process. & System Theory, Faculty of Engineering, CAU, 24143 Kiel, E-Mail: uh@tf.uni-kiel.de

Abstract

An acoustic-tube model of a speaker's vocal tract can be used for observations of the speaker's articulation. This is of interest here in order to gain insight into articulatory weaknesses of Parkinson's-Disease (PD) patients, who suffer from speaking problems: Beside the common evaluations of auditory and some instrumental parameters, a direct observation of vocal-tract movements can be an additional diagnostic aid. Due to its simplicity, the direct estimation of vocal-tract areas from an LPC analysis is used here, despite its well-known impreciseness. The usefulness in general was shown in a pre-study; now, detailed results of an application on sustained vowels as well as fluent speech from healthy, slightly handicapped, and strongly impaired speakers are presented. Examples of short-time and average shapes are shown and compared, and various parameters are derived, which condense the observations and differences into numerical values. These terms cover largely overlapping ranges for patients in different disease states; so, they are not to be used for any single-value classification of the dysarthric severeness, but offered to the medical personnel together with the varying shapes in order to augment logopedic diagnosis and therapy.

Introduction

Within the usual recursive computation of LPC coefficients [1], the key terms used in each step can be interpreted as reflection coefficients describing area changes in a *much* simplified vocal-tract model [2]. The model consists of (N+1) circular tubes, each with an individual, time-varying cross-sectional area $q_i(k)$and identical length $l/(N+1)$, where l is the total tract length and N is the order of the linear predictor to be designed. This approach – despite its known inaccuracies (see, e.g., [3, 4]) – is now applied in its simplest form as a fast observation tool for dysarthric speech signals.

In [5], the approach was checked for its general applicability. It was found that the model is not sufficient to display actions of single anatomical parts of the vocal tract, but it does give information about articulatory activity and differences. Thus, a detailed analysis of dysarthric speech and a comparison both with healthy speakers' signals and with those of patients in different disease and treatment situations is worth an effort.

In all cases, the acoustic signals were sampled with a rate $f_s = 8kHz$. With an LPC order N = 8, the time varying tube areas $q_i(k), i \in \{0,1,\ldots,8\}$ were calculated for n_{fr} 50%-overlapping signal frames of length 128 ms, starting with $q_0(k) \equiv Q_0 = 3.14cm^2$, i.e., the simple assumption of a constant tube radius of 1 cm at the glottis.

Vowels of Healthy Speakers

In Fig. 3 of [5], the analysis result for a healthy speaker's vowel /*a:*/ was depicted, with the intermediate results of signal, spectrum, reflection coefficients, time-averaged areas

$$Q_i = {}^1\!/_{n_{fr}} \cdot \sum_{k=1}^{n_{fr}} q_i(k) \tag{1}$$

and radii $R_i = \sqrt{Q_i/\pi}$. In Fig. 1, the "shape" described by $\pm R_i$ is repeated, together with 2 other variants of this vowel. The shape variability is indicated by thin lines $+\sigma_{Q_i}$ above and $-\sigma_{Q_i}$ below the average curves; the standard deviation σ_{Q_i} follows from the variance

$$\sigma_{Q_i}^2 = {}^1\!/_{n_{fr}} \cdot \sum_{k=1}^{n_{fr}} [q_i(k) - Q_i]^2. \tag{2}$$

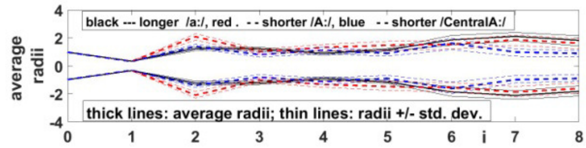

Fig. 1: 3 /*a:*/ variants with average radii and radius variations by $\pm$ standard deviation σ_{Q_i}.

The above-named variation (see also [5]) is clarified in Fig. 2: The shape of a *sustained* vowel although spoken by a *healthy* person is not at all constant, also visible in Fig. 3, for 4 speakers, as a time-dependent mean opening

$$r(k) = \sqrt{q\,(k)/\pi}, \tag{3a}$$

with

$$q(k) = {}^1\!/_{(N-1)} \cdot \sum_{i=2}^{N} q_i(k) \tag{3b}$$

(dropping the almost constant, dominant terms at i=0,1).

Fig. 4 depicts, linearly interpolated, the time-averaged area functions Q_i of 4 speakers and their variances. There are obvious differences even between healthy speakers. But the shape is more or less consistent and non-monotone, the variance is smaller for small mean areas [1].

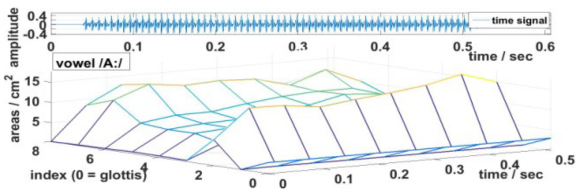

Fig. 2: Time behaviour of the areas for the vowel /*A:*/.

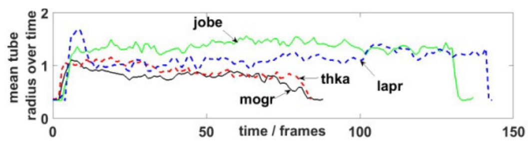

Fig. 3: Time dependence of mean tube sizes $r(k)$ for 4 healthy speakers.

Patients' Vowels

The above descriptions have now to be compared to the

[1] The "large" values in Fig. 4 are not too dramatic: Q_i = 12 means R_i = 1.95 cm, changed by a variance 25 to $\sqrt{(12 \pm 5)/\pi} = 2.3\,cm \vee 1.5cm$.

corresponding ones for PD patients.

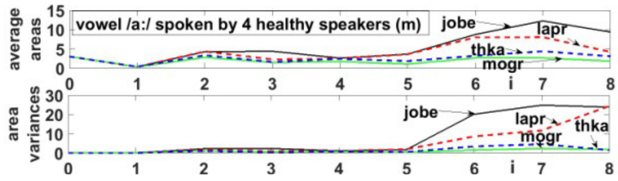

Fig. 4: Average areas and variances for 4 healthy speakers.

Patients with Little Speaking Handicaps

Patients regarded in this section have no severe speaking deficits, with NTID values between 4 and 5 [2]. As Fig. 8 in [5], Figs. 5 and 6 say that, nevertheless, these patients tend to have rather small tube openings and (therefore) low variances: "kare" represents a healthy speaker (f) with relatively small values, and her curves are (mainly) above those of the patients. But the patients show strong individual differences, and there are even some cases with larger areas and variances (see Fig. 9 in [5]). So, the "kare" curve cannot serve as a simple classifier boundary between "healthy" and "slightly dysarthric".

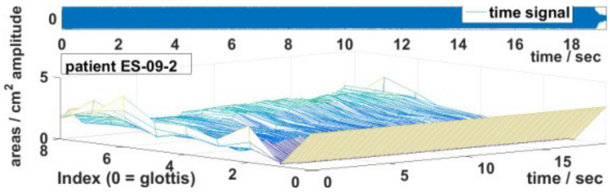

Fig. 5: Time behaviour of the vowel /*a*:/ of a patient with small handicap.

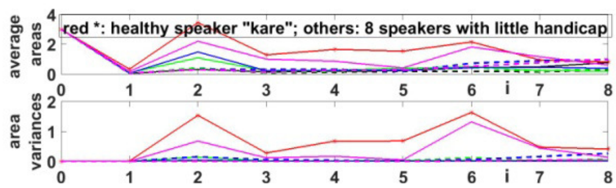

Fig. 6: Areas and variances for /*a*:/ from a healthy speaker and 8 patients with small handicap.

In Fig. 7 the mostly small openings reappear as low mean radii for 6 patients, smaller than those of the healthy persons in Fig. 3, and small are their variations over time.

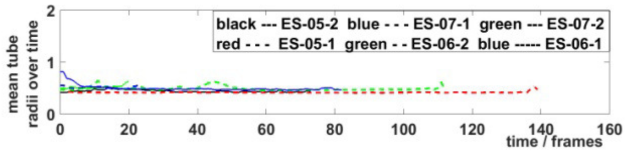

Fig. 7: Time dependence of mean tube sizes for 6 patients with small problems.

Patients with Strong Speaking Handicaps

Patients with NTID values around 2 were examined next. Fig. 12 in [5] indicates that they may have average areas and variances *below or above* those of a weakly articulating healthy person. Fig. 8 now shows that even values of a well-articulated healthy person's /*a:*/ may in parts be surpassed by that of a poorly speaking patient [3]. Again, individual behaviours are very different, no simple limits are possible. The background becomes visible, if the time variation of the model tube is examined: As to be seen in Fig. 9, there are widely different variants of speaking styles among handicapped persons – there are "decaying" (and also "growing"), "fluctuating", or "burst-like" articulation efforts. Fig. 10 displays these variants in terms of 4 time-varying mean tube openings. They show that these patients may have a very small tube diameter all the time or exhibit larger, but very instationary radii.

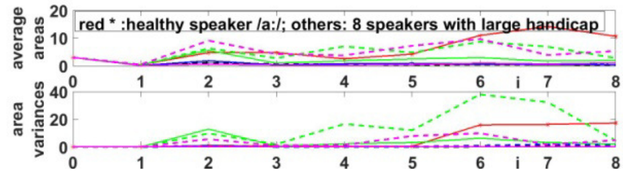

Fig. 8: Average areas and variances from a healthy speaker and 8 heavily handicapped persons.

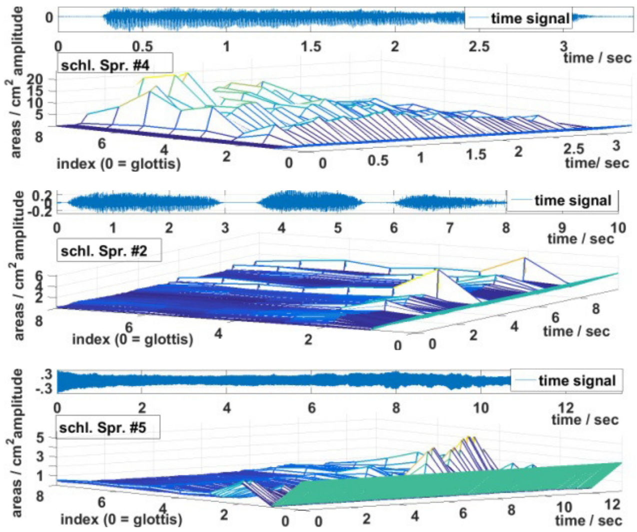

Fig. 9: Largely differing time behaviours of the tube model for 3 strongly-handicapped persons.

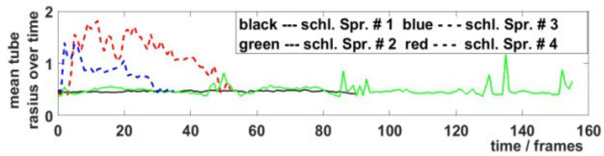

Fig. 10: Time dependence of mean tube sizes for 4 strongly impaired speakers.

Healthy Speakers' Fluent Speech

Fluent speech contains *varying* sounds. So, the short-time view of the tube model should reflect this by time variations – as seen in Fig. 11. So, from an averaged area, only an information about a larger or smaller mean opening can be expected – of interest are its time-dependent form $r(k)$ as in Figs. 3, 7, and 10, and the averaged variances: Both show the ability to *open* and to *move* the articulators, in contrast to the vowel case where a low variation is positive. Fig. 12 depicts the results for the healthy person of Fig. 11, for a comparison with patients below.

Patients' Fluent Speech

Figs. 13 and 14 show that, indeed, patients produce a much reduced activity, visible in the short-time plots as well as via the averaged variances and time-varying radii.

[2]The modified "National Technical Institute of the Deaf" (NTID) scale defines a value "6" for perfectly intelligible,"1" for unintelligible speech.

[3] German annotation in the figures: "schl. Spr." = poor speech / strongly handicapped speaker.

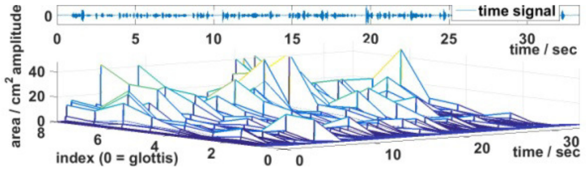

Fig. 11: Fluent-speech tube model for a healthy speaker.

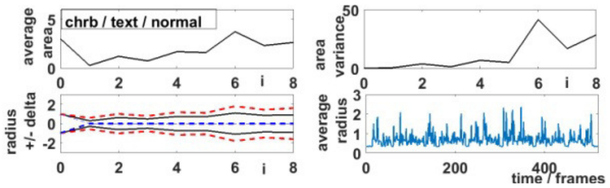

Fig. 12: Analysis of the speech signal in Fig. 11.

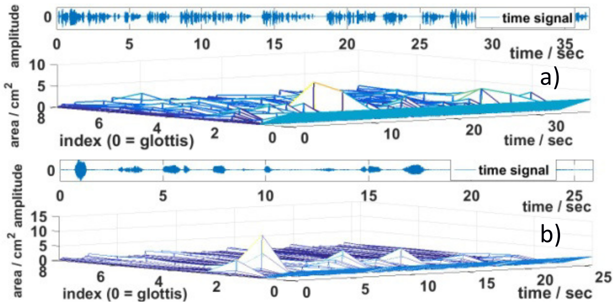

Fig. 13: Short-time tube behaviours for a patient with a) small and b) strong speaking problems.

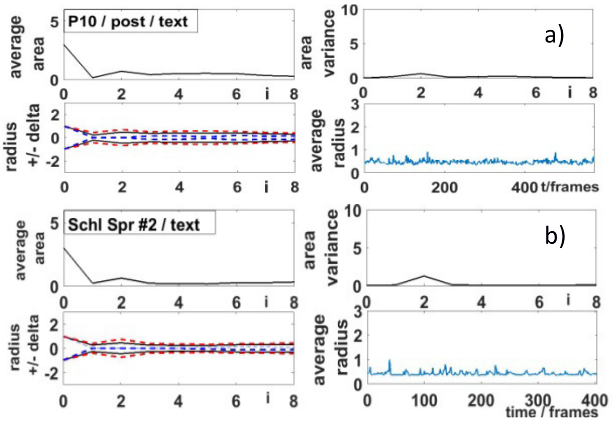

Fig. 14: Analyses of the speech signals in Fig. 13.

Parametric Evaluation

Beside the above graphics, a "condensed" information may help in diagnosis, as given by the following parameters (see definitions in (1-3)):

$$\mu_Q = \frac{1}{n_{fr}\cdot(N-1)}\cdot\sum_{i=2}^{N}\sum_{k=1}^{n_{fr}} q_i(k) = \frac{1}{N-1}\cdot\sum_{i=2}^{N} Q_i, \quad (4)$$

$$\sigma_Q^2 = \frac{1}{N-1}\cdot\sum_{i=2}^{N}(Q_i-\mu_Q)^2 \quad (5)$$

and

$$\mu_{\sigma_Q^2} = \frac{1}{n_{fr}\cdot(N-1)}\cdot\sum_{i=2}^{N}\sum_{k=1}^{n_{fr}}(q_i(k)-Q_i)^2 : \quad (6)$$

- Both for "good" vowels and fluent speech, the "overall-mean tube size "μ_Q should be "large" (e.g., > Q_0),
- as well as the "mean shaping" σ_Q^2 along the tube length;
- the "mean instationarity" $\mu_{\sigma_Q^2}$ should be small for vowels (smaller than σ_Q^2), but large for fluent speech. The distribution of these parameters for a number of test persons, first for sustained vowels, is shown in Fig. 15.

Here, certain trends are visible: Healthy persons cover a large range of mean openings and longitudinal variation, much larger than patients with a beginning dysarthria. However, heavily handicapped patients may behave similarly. The same holds for the time variation. The clearest indication for a problem is found in the quotient $\mu_{\sigma_Q^2}/\sigma_Q^2$: Time versus longitudinal variation grows with increasing speaking difficulties.

For fluent speech, the corresponding Fig. 16 gives clearer indications: All parameters decay with increasing speaking

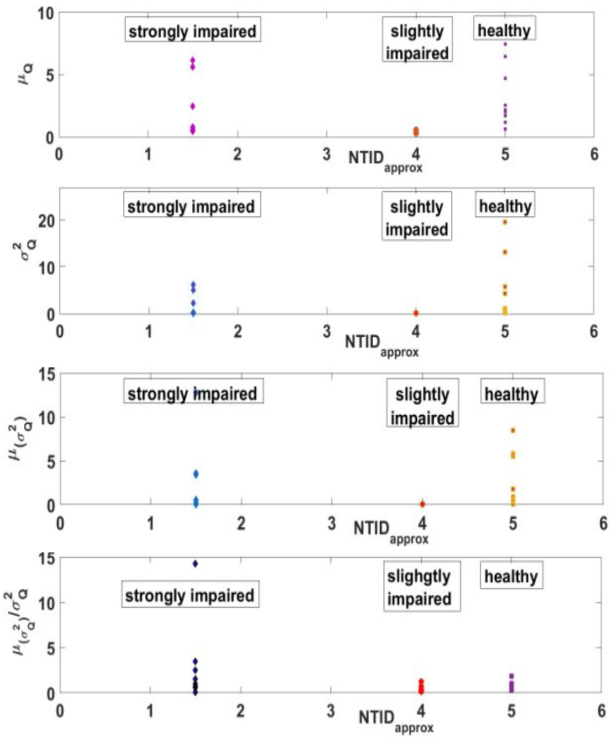

Fig. 15: Parameters in different health states: Vowels.

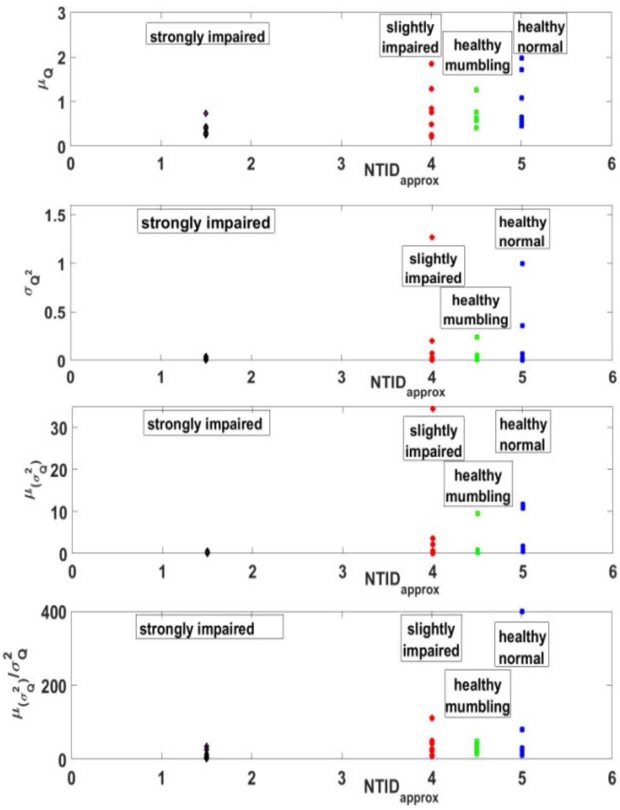

Fig 16: Parameters in different health states: Read text.

problems – though only *on average*: There is no threshold clearly separating the groups. A clearer *classification* can be based on multiple parameters, e. g. model parameters as used here or others, and a Gaussian mixture or neural-network approach [6].

Patients in Varied Treatment States

A reason for the wide variation of results for patients with the same problem severity and the subsequent overlap in Figs. 15 and 16 is the individually quite different behaviour of – also healthy – persons. This individuality can, however, be possibly overcome, if the above techniques are applied to identical persons in different treatment states.

A number of patients in the data base underwent a (surgical or drug-based) treatment alleviating bodily symptoms. A first tube-model comparison of a patient's vowel is already included in Fig. 9 of [5]: The "post-treatment" opening is larger, the variance is smaller than in the "pre-treatment" description; so, a success may be assumed also for the communication ability. Fig. 7, however, says that this is no general result: The curves describing cases with identifications "ES-xx-2" belong to the same persons after treatment as "ES-xx-1" before, and there is no clear result. But this is no real surprise: It was observed before that bodily enhancements do not always coincide with speaking improvements [7, 8]. A better way towards an enhanced intelligibility is a training of a loud and clear speaking style, like, e.g., the "Lee-Silverman Voice Training" [9].

Patients before and after Training

Vowels of 4 patients and the healthy (low-articulation) speaker "baka" are compared in Fig 17: Full lines indicate the speaking style after the training; they are always above the dashed ones describing the pre-training abilities. Also, "baka"-tube values are surpassed. The – unwanted – variances increase as well; but Fig. 18 clarifies that their sizes grow to "normal" values without too much disturbance, compared to Figs. 3 or 9. Ongoing work concerns the verification and further evaluation for such patients.

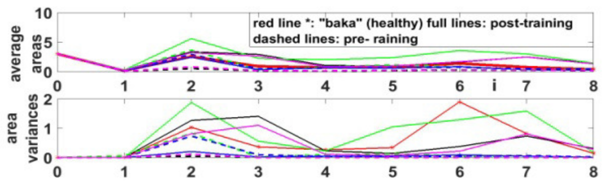

Fig. 17: Average areas and variances from healthy speaker "baka" and 4 persons before and after a voice training.

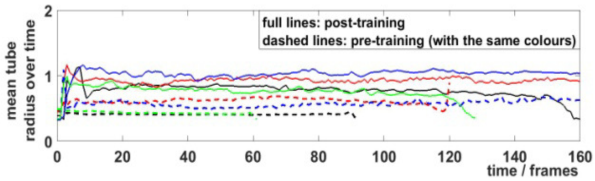

Fig. 18: Time-varying mean radii for 4 patients before and after a voice training.

Conclusions and Outlook

The simple LPC-based acoustic-tube model for the vocal tract was applied to sustained vowels and fluent speech of healthy speakers and dysarthric PD patients. Details of articulation differences became visible, which, together with some parametric model descriptions, promise to be helpful for the logopedic diagnosis. Such a use is under investigation, together with a verification for more patients and possibly an enhanced, but still simple model.

Acknowledgements

The basic MATLAB program for the LPC analysis was provided by G. Schmidt, the data base by A. Nebel, the extracted speech files by C. Baasch. All of them contributed to many discussions. The author wishes to express his gratitude to them. Sincere thanks are due to DFG (German Research Foundation) for Grant HE 4465/6-1.

References

[1] Markel, J.D., Gray, A.H.: Linear prediction of Speech. Springer, New York, 1976

[2] Wakita, H.: Direct Estimation of the Vocal-Tract Shape by Inverse Filtering of Acoustic Speech Waveforms. IEEE Trans. Audio El. Acoust. AU-21 (1973), 417-427

[3] Wakita, H.: Estimation of Vocal-Tract Shapes from Acoustical Analysis of the Speech Wave: The State of the Art. IEEE Trans. Acoust., Speech, Sig. Proc. ASSP-27 (1979), 281-285

[4] Deng, H., Beddoes, M.P., Ward, R.K., Hodgson, M.: Obtaining the Vocal-Tract Area Function from the Vowel Sound. Canad. Acoust. 31 (2003), 40-41

[5] Heute, U., Schmidt, G.: Observations from a Simple Vocal-Tract-Model's Behaviour for PD-Dysarthric Speech: Applicability. ITG Conf. Speech Comm., Oldenburg, 2018.

[6] Baasch, C.: Instrumentelle Analyse von Parkinson-Sprache. Diss. (in German), DSS/EIT/TF/CAU, Kiel, 2018.

[7] Xie, Y., Zhang, Y., Zjeng, Z., Liu, A., Wang, X., Zhuang, P., Li, Y., Wang, X.: Changes in Speech Characters of Patients with Parkinson's Disease after Bilateral Subthalamic-Nucleus Stimulation. J. Voice 25 (2011), 751-758

[8] Sauvageau, V.M., Roy,J.-P., Cantin, L., Prud'Homme, M., Langlois, M., Macoir, J.: Articulatory Changes in Vowel Production Following Deep-Brain Stimulation and Levodopa Intake in Parkinson's Disease. Park. Dis. (2015), Art. ID 382320

[9] Ramig, L.A., Mead, C.L., deSanto, L.: Voice Therapy and Parkinson's Disease. American Speech & Hearing Association 30 no. 109 (1988), 128