Universität Rostock Traditio et Innovatio

Chair of Mechatronics

Interval Analysis for the Representation of Phoneme Databases in Speech Recognition Systems: Fundamentals of a Computer-Based Assistance System in Speech Therapy

swim2016: Summer Workshop on Interval Methods 2016

Lyon, France

June 21st, 2016

Andreas Rauh[1], Susann Tiede[2], Cornelia Klenke[2]

[1]Chair of Mechatronics, University of Rostock, Germany
[2]Speech Therapists, Altentreptow, Germany

Contents

- Classification of language disorders and project aims
- Characteristic properties of voiced and unvoiced phonemes in speech signals
- Approaches for frequency estimation
  - Offline short-time Fourier analysis
  - Observer-/ Filter-based approaches
- Phoneme-based segmentation of speech signals
- Interval-based feature representation
- Preliminary classification results
  - Distinction between voiced and unvoiced sounds (based on thresholds for the estimated standard deviations)
  - Distinction between different voiced sounds (vowels, based on the estimated narrow-band frequencies)
- Conclusions and outlook on future work

Development of a Computer-Based Assistance System in Speech Therapy (1)

Classification of language disorders

- Pronunciation
- Grammar
- Lexicon

Project aims

1. Automatic transcription and preprocessing of spoken text involving erroneous pronunciation
2. Automatic classification of pronunciation disorders
3. Grammatical analysis of freely spoken language

Development of a Computer-Based Assistance System in Speech Therapy (2)

Classification of language disorders

- Pronunciation
- Grammar
- Lexicon

Note

- State-of-the-art speech recognition and text processing systems replace *incorrect* items by *seemingly* correct ones
- These substitutions prevent the classification of language disorders

Development of a Computer-Based Assistance System in Speech Therapy (3)

Classification of language disorders

- Pronunciation
- Grammar
- Lexicon

Note

- Requirement to develop a new estimation and classification scheme for speech signals
- Tedious and time-consuming work of speech therapists if freely spoken language is to be analyzed (in contrast to standardized test procedures) $\Longrightarrow$ Need for repeated listening of recorded speech

Characteristic Properties of Speech Signals (1)

Distinction between voiced and unvoiced phonemes

- Phonemes as the basic building block of syllables, from which words are formed
- Characterized by specific features
  - Value of basis frequency is speaker dependent
  - Format frequencies (basis frequency + higher frequency values)
  - Higher formant frequencies are typically not integer multiples of the basis frequency
  - Bandwidth of included frequency ranges varies for voiced/ unvoiced phonemes
- Voiced phonemes: E.g. *normal vowels*, characterized by several **sharp formant frequencies**
- Unvoiced phonemes: E.g. *whispered vowels* and *fricatives* such as ch, ss, sh, f, characterized by **wide blurred formant frequency ranges**

Characteristic Properties of Speech Signals (2)

Mechanism of sound production

- Voiced phonemes
  - Produced by vibrations of the vocal folds
  - Vocal folds represent a fluidic resistance against the outflow of air expelled from the lungs
- Unvoiced phonemes
  - Turbulent, partially irregular, air flow
  - Negligible vibrations of the vocal folds
  - Fizzing sounds
  - Originating between teeth and lips as well as between tongue and hard/ soft palate

Characteristic Properties of Speech Signals (3)

State-of-the-art speech recognition systems

- Use of offline frequency analysis
  1. Cut the sound sequence into short temporal slices of typically $10 - 50\,\mathrm{ms}$ length
  2. Perform a short-time Fourier analysis for each of these time slices (partly with overlapping time windows)
  3. Determine a measure of similarity with phoneme-dependent frequency spectra (usually by the application of cross-correlation functions in the frequency domain)

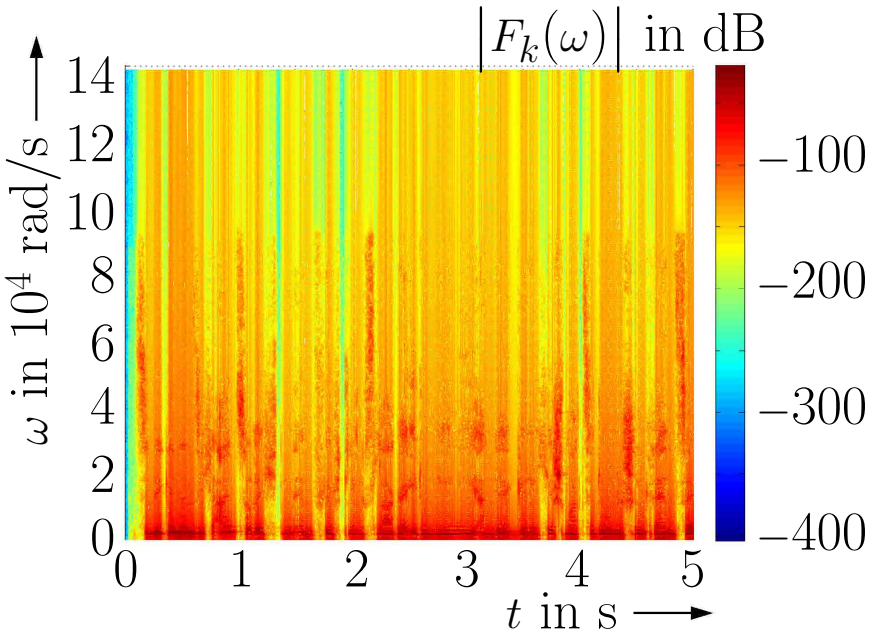
Fourier Analysis of a Benchmark Dataset (1)

Parameterization of the DFT analysis (sampling frequency $f_\mathrm{s} = 44.1\,\mathrm{kHz}$)

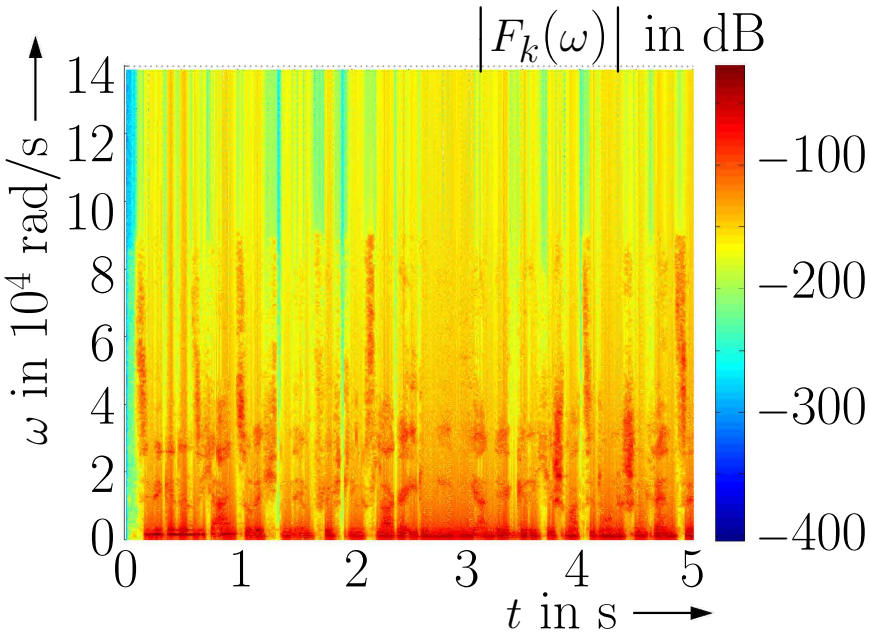
	DFT 1	DFT 2	DFT 3
Length of time window in samples (N)	512	1024	2048
Length of time window in ms	11.6	23.2	46.4
Sample shift between two DFT evaluations	64	64	64
Spectral resolution Δf in Hz	86.5	43.2	21.6

Fourier Analysis of a Benchmark Dataset (2)

Output for the setting DFT 1



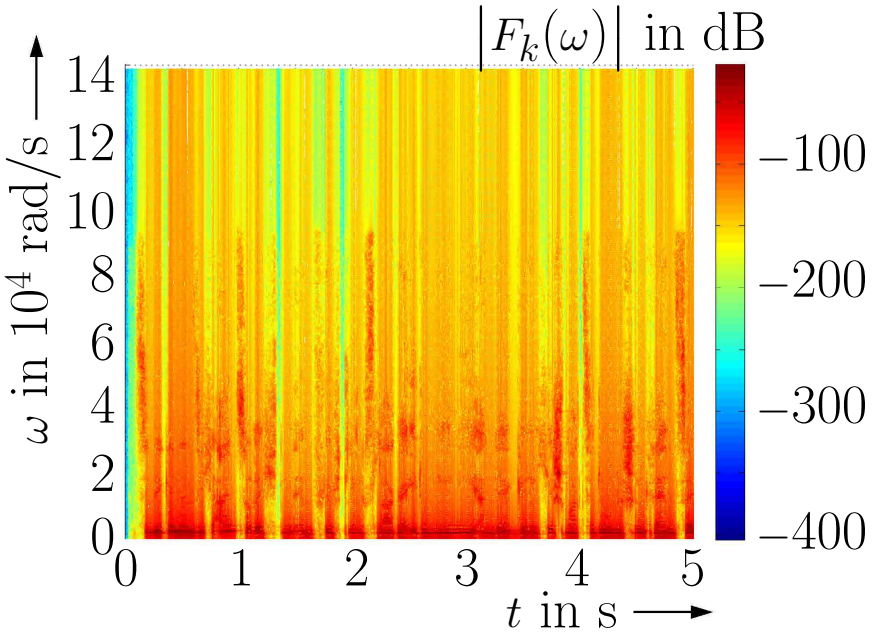
Output for the setting DFT 2



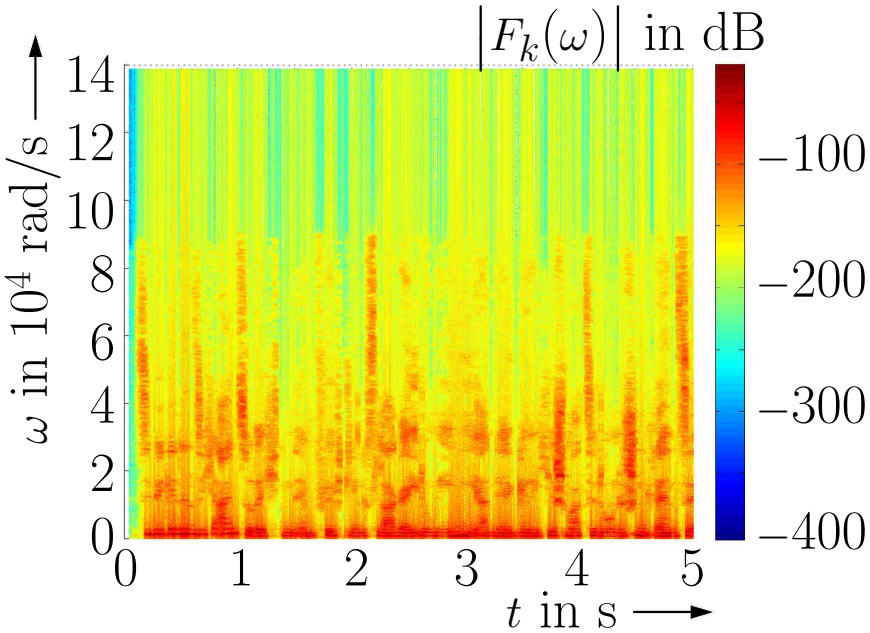
Increasing the number of sampling points enhances the detectability of formant frequencies

Fourier Analysis of a Benchmark Dataset (3)

Output for the setting DFT 1



Output for the setting DFT 3



Points of time with significant changes in the spectrum represent candidates for transition between subsequent phonemes

Filter- and Observer-Based Online Frequency Analysis (1)

Signal model

Measured speech signal is approximated by a superposition of different harmonic components

$$y_{\mathrm{m}}(t) \approx y_{\mathrm{m},n}(t) = \sum_{i=1}^{n} \left(\alpha_i \cdot \cos\left(\omega_i \cdot t + \phi_i\right)\right)$$

with the basis frequency $\omega_1 > 0$, further harmonic signal components $\omega_2, \ldots, \omega_n$, $\omega_{i+1} > \omega_i$, $i \in \mathbb{N}$, the signal amplitudes α_i, and the phase shifts ϕ_i

Filter- and Observer-Based Online Frequency Analysis (2)

State-space representation of the i-th frequency component

Continuous-time system model

$$\begin{aligned}\dot{x}_{3i-2}(t) &= -x_{3i}(t) \cdot \alpha_i \cdot \sin\left(x_{3i}(t) \cdot t + \phi_i\right) =: x_{3i-1}(t)\\ \dot{x}_{3i-1}(t) &= -x_{3i}^2(t) \cdot \alpha_i \cdot \cos\left(x_{3i}(t) \cdot t + \phi_i\right)\\ \dot{x}_{3i}(t) &= 0\end{aligned}$$

Compact matrix-vector notation

Continuous-time system model

$$\dot{\mathbf{x}}_i(t) = \mathbb{A}_i\left(x_{3i}(t)\right) \cdot \mathbf{x}_i(t)\ , \quad \mathbf{x}_i(t) = \begin{bmatrix} x_{3i-2}(t) \\ x_{3i-1}(t) \\ x_{3i}(t) \end{bmatrix}$$

Filter- and Observer-Based Online Frequency Analysis (2)

State-space representation of the i-th frequency component

Continuous-time system model

$$\begin{aligned}\dot{x}_{3i-2}(t) &= -x_{3i}(t) \cdot \alpha_i \cdot \sin\left(x_{3i}(t) \cdot t + \phi_i\right) =: x_{3i-1}(t)\\ \dot{x}_{3i-1}(t) &= -x_{3i}^2(t) \cdot \alpha_i \cdot \cos\left(x_{3i}(t) \cdot t + \phi_i\right)\\ \dot{x}_{3i}(t) &= 0\end{aligned}$$

Compact matrix-vector notation

Continuous-time system model

$$\dot{\mathbf{x}}_i(t) = \mathbb{A}_i\left(x_{3i}(t)\right) \cdot \mathbf{x}_i(t)\ , \quad \mathbb{A}_i\left(x_{3i}(t)\right) = \left[\begin{array}{cc:c} 0 & 1 & 0 \\ -x_{3i}^2(t) & 0 & 0 \\ \hdashline 0 & 0 & 0 \end{array}\right]$$

Filter- and Observer-Based Online Frequency Analysis (2)

State-space representation of the i-th frequency component

Continuous-time system model

$$\begin{aligned}\dot{x}_{3i-2}(t) &= -x_{3i}(t) \cdot \alpha_i \cdot \sin\left(x_{3i}(t) \cdot t + \phi_i\right) =: x_{3i-1}(t)\\ \dot{x}_{3i-1}(t) &= -x_{3i}^2(t) \cdot \alpha_i \cdot \cos\left(x_{3i}(t) \cdot t + \phi_i\right)\\ \dot{x}_{3i}(t) &= 0\end{aligned}$$

Compact matrix-vector notation

Continuous-time system model

$$\dot{\mathbf{x}}_i(t) = \mathbb{A}_i\left(x_{3i}(t)\right) \cdot \mathbf{x}_i(t)\ , \quad y_i(t) = \check{\mathbf{c}}_i^T \cdot \mathbf{x}_i(t) \quad \text{with} \quad \check{\mathbf{c}}_i^T = \begin{bmatrix}1 & 0 & 0\end{bmatrix}$$

Filter- and Observer-Based Online Frequency Analysis (2)

State-space representation of the i-th frequency component

Continuous-time system model

$$\begin{aligned}
\dot{x}_{3i-2}(t) &= -x_{3i}(t) \cdot \alpha_i \cdot \sin\left(x_{3i}(t) \cdot t + \phi_i\right) =: x_{3i-1}(t) \\
\dot{x}_{3i-1}(t) &= -x_{3i}^2(t) \cdot \alpha_i \cdot \cos\left(x_{3i}(t) \cdot t + \phi_i\right) \\
\dot{x}_{3i}(t) &= 0
\end{aligned}$$

Concatenation of all models $i = 1, \ldots, n$

Continuous-time system model

$$\dot{\mathbf{x}}(t) = \mathbf{A}\left(\mathbf{x}(t)\right) \cdot \mathbf{x}(t) \quad \text{with} \quad \mathbf{x}(t) \in \mathbb{R}^{3n}$$

Filter- and Observer-Based Online Frequency Analysis (2)

State-space representation of the i-th frequency component

Continuous-time system model

$$\begin{aligned}\dot{x}_{3i-2}(t) &= -x_{3i}(t) \cdot \alpha_i \cdot \sin\left(x_{3i}(t) \cdot t + \phi_i\right) =: x_{3i-1}(t)\\ \dot{x}_{3i-1}(t) &= -x_{3i}^2(t) \cdot \alpha_i \cdot \cos\left(x_{3i}(t) \cdot t + \phi_i\right)\\ \dot{x}_{3i}(t) &= 0\end{aligned}$$

Concatenation of all models $i = 1, \ldots, n$

Continuous-time system model

$$\mathbf{A}\left(\mathbf{x}(t)\right) = \begin{bmatrix} \mathbb{A}_1\left(x_3(t)\right) & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbb{A}_2\left(x_6(t)\right) & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbb{A}_n\left(x_{3n}(t)\right) \end{bmatrix}$$

Filter- and Observer-Based Online Frequency Analysis (2)

State-space representation of the i-th frequency component

Continuous-time system model

$$\begin{aligned}\dot{x}_{3i-2}(t) &= -x_{3i}(t) \cdot \alpha_i \cdot \sin\left(x_{3i}(t) \cdot t + \phi_i\right) =: x_{3i-1}(t)\\ \dot{x}_{3i-1}(t) &= -x_{3i}^2(t) \cdot \alpha_i \cdot \cos\left(x_{3i}(t) \cdot t + \phi_i\right)\\ \dot{x}_{3i}(t) &= 0\end{aligned}$$

Concatenation of all models $i = 1, \ldots, n$

Continuous-time system model

$$y_{\mathrm{m},n}(t) = \mathbf{c}^T \cdot \mathbf{x}(t) \quad \text{with} \quad \mathbf{c}^T = \begin{bmatrix}\check{\mathbf{c}}_1^T & \check{\mathbf{c}}_2^T & \ldots & \check{\mathbf{c}}_n^T\end{bmatrix}$$

Filter- and Observer-Based Online Frequency Analysis (2)

State-space representation of the i-th frequency component

Continuous-time system model

$$\begin{aligned}\dot{x}_{3i-2}(t) &= -x_{3i}(t)\cdot\alpha_i\cdot\sin\left(x_{3i}(t)\cdot t+\phi_i\right) =: x_{3i-1}(t)\\ \dot{x}_{3i-1}(t) &= -x_{3i}^2(t)\cdot\alpha_i\cdot\cos\left(x_{3i}(t)\cdot t+\phi_i\right)\\ \dot{x}_{3i}(t) &= 0\end{aligned}$$

Compact matrix-vector notation

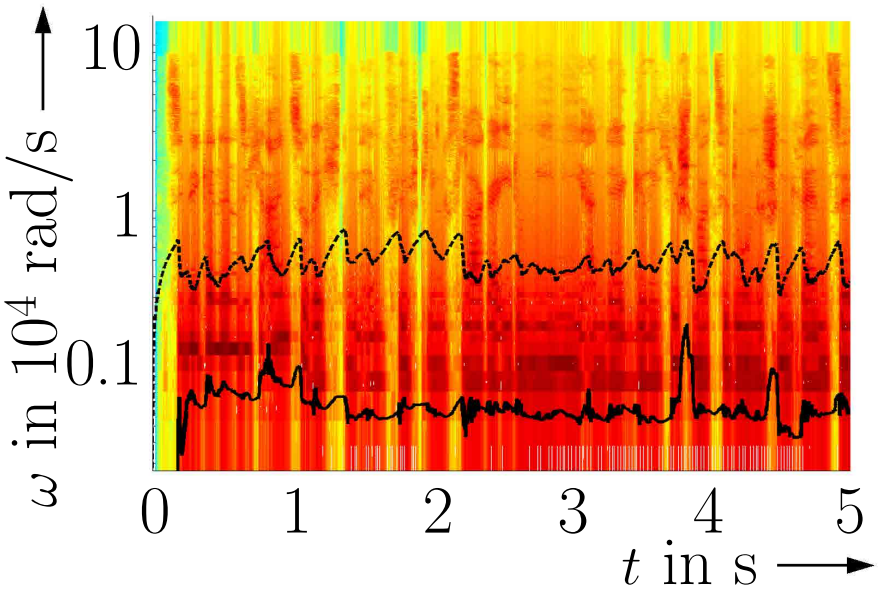
Discrete-time notation for an Extended Kalman Filter design

$$\mathbf{x}_{k+1}=\exp\left(T_\mathrm{s}\cdot\mathbf{A}\left(\mathbf{x}_k\right)\right)\cdot\mathbf{x}_k+\mathbf{w}_k =: \mathbf{A}_k^\mathrm{d}\cdot\mathbf{x}_k+\mathbf{w}_k,\ f_{w,k}\left(\mathbf{w}_k\right)=\mathcal{N}(\boldsymbol{\mu}_{w,k},\mathbb{C}_{w,k})$$

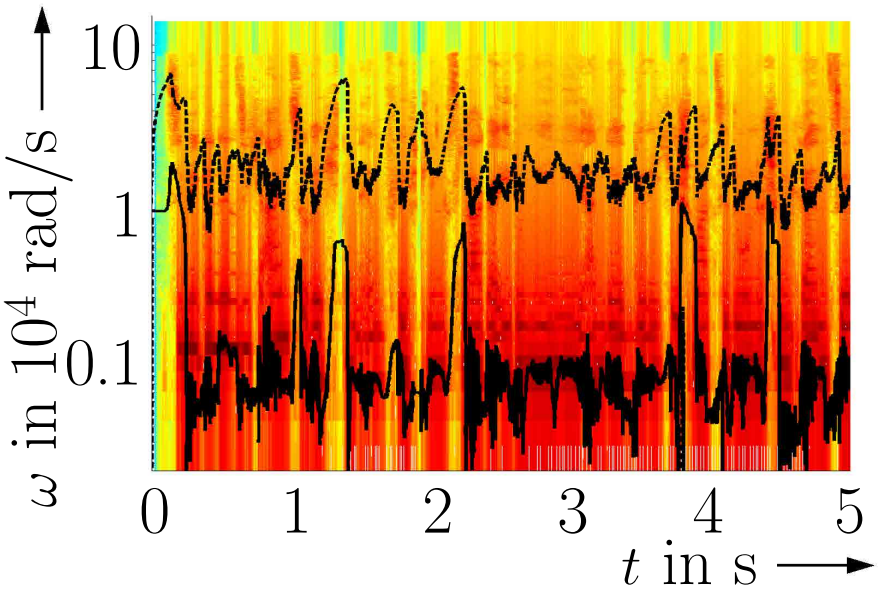
$$y_k = y_{\mathrm{m},n,k} := y_{\mathrm{m},n}(t_k) = \mathbf{c}^T\cdot\mathbf{x}_k+v_k\ ,\quad f_{v,k}\left(v_k\right)=\mathcal{N}\left(\mu_{v,k},\mathbb{C}_{v,k}\right)$$

Filter- and Observer-Based Online Frequency Analysis (3)

Estimate for ω_1



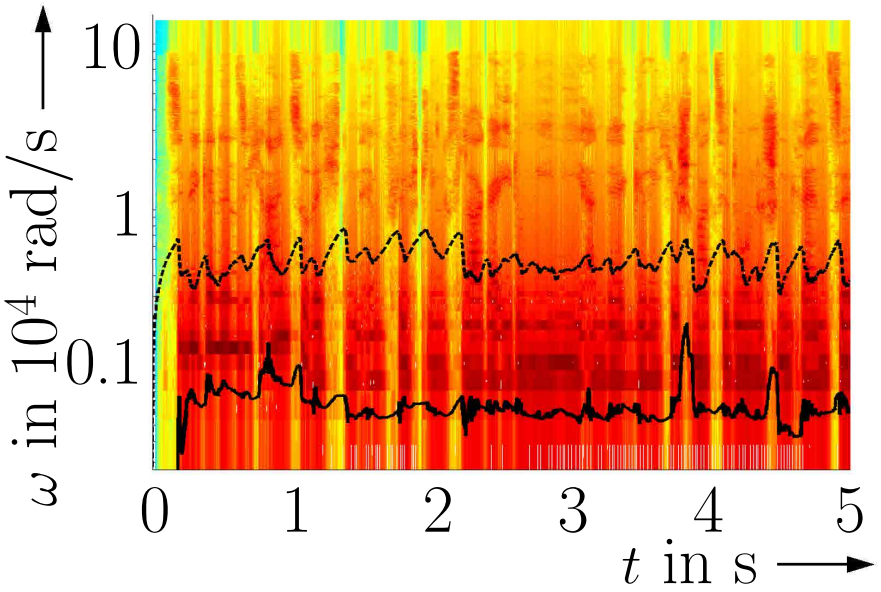
Estimate for ω_2



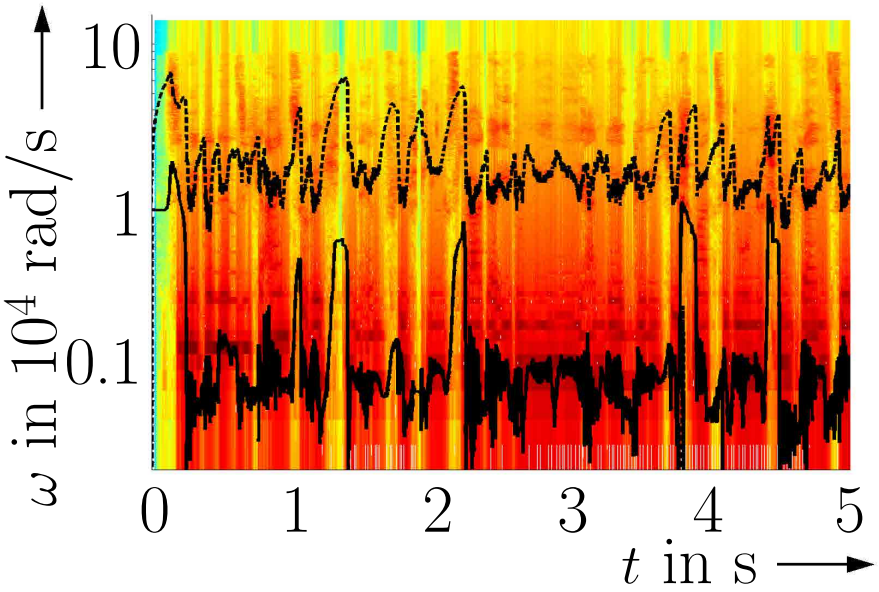
Estimation results for $n = 2$: expected value $\mu^e_{x,k,3i}$ (solid lines) and upper frequency bound $\mu^e_{x,k,3i} + 3\left(\sqrt{\mathbb{C}^e_{x,k,(3i,3i)}}\right)$; $n = 3$ would be required to estimate the next formant frequency ω_3

Filter- and Observer-Based Online Frequency Analysis (3)

Estimate for ω_1



Estimate for ω_2



A. Rauh, S. Tiede, and C. Klenke. *Observer and Filter Approaches for the Frequency Analysis of Speech Signals.* **and** *Stochastic Filter Approaches for a Phoneme-Based Segmentation of Speech Signals.* In Proc. of 21st IEEE Intl. Conference on Methods and Models in Automation and Robotics MMAR 2016, Miedzyzdroje, Poland, 2016.

Phoneme-Based Segmentation of the Speech Signal (1)

Normalization of the Extended Kalman Filter outputs (L samples)

- *Normalized expectation* of the i-th formant frequency

$$\tilde{\mu}^{e}_{x,k,3i} = \frac{\mu^{e}_{x,k,3i}}{\frac{1}{L}\sum_{k=1}^{L} \mu^{e}_{x,k,3i}}$$

- *Normalized (co-)variance* of the i-th formant frequency

$$\tilde{\mathbb{C}}^{e}_{x,k,(3i,3i)} = \frac{\mathbb{C}^{e}_{x,k,(3i,3i)}}{\frac{1}{L}\sum_{k=1}^{L} \mathbb{C}^{e}_{x,k,(3i,3i)}}$$

Phoneme-Based Segmentation of the Speech Signal (2)

Variation rate classification concerning the expectations of the estimated formant frequencies

- Absolute variation rate

$$\Delta\tilde{\mu}^{e}_{x,k,3i} = \left|\tilde{\mu}^{e}_{x,k+1,3i} - \tilde{\mu}^{e}_{x,k,3i}\right|$$

- Candidates for phoneme boundaries are characterized by

$$\Delta\tilde{\mu}^{e}_{x,k,3i} > \overline{\Delta\tilde{\mu}}$$

Note

Temporal distance between two subsequent boundaries needs to be larger than $\Delta T_{\min}$

Phoneme-Based Segmentation of the Speech Signal (3)

Variation rate classification concerning the (co-)variances of the estimated formant frequencies

- Absolute variation rate

$$\Delta\tilde{\mathbb{C}}^{e}_{x,k,(3i,3i)} = \left|\tilde{\mathbb{C}}^{e}_{x,k+1,(3i,3i)} - \tilde{\mathbb{C}}^{e}_{x,k,(3i,3i)}\right|$$

- Candidates for phoneme boundaries are characterized by

$$\Delta\tilde{\mathbb{C}}^{e}_{x,k,(3i,3i)} > \overline{\Delta\tilde{\mathbb{C}}}$$

Note

Temporal distance between two subsequent boundaries needs to be larger than $\Delta T_{\min}$

Phoneme-Based Segmentation of the Speech Signal (3)

Variation rate classification concerning the (co-)variances of the estimated formant frequencies

- Absolute variation rate

$$\Delta\tilde{\mathbb{C}}^{e}_{x,k,(3i,3i)} = \left|\tilde{\mathbb{C}}^{e}_{x,k+1,(3i,3i)} - \tilde{\mathbb{C}}^{e}_{x,k,(3i,3i)}\right|$$

- Candidates for phoneme boundaries are characterized by

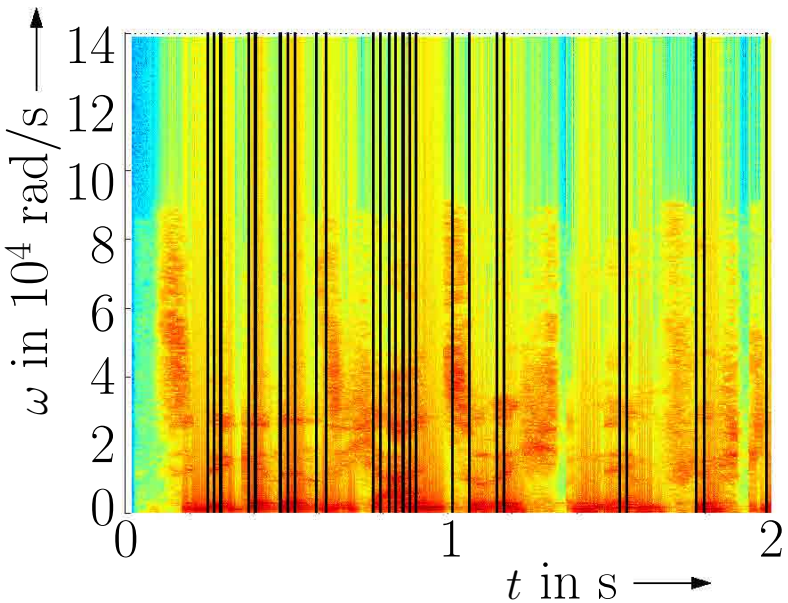
$$\Delta\tilde{\mathbb{C}}^{e}_{x,k,(3i,3i)} > \overline{\Delta\tilde{\mathbb{C}}}$$

Note: Combinations of both classifiers are possible

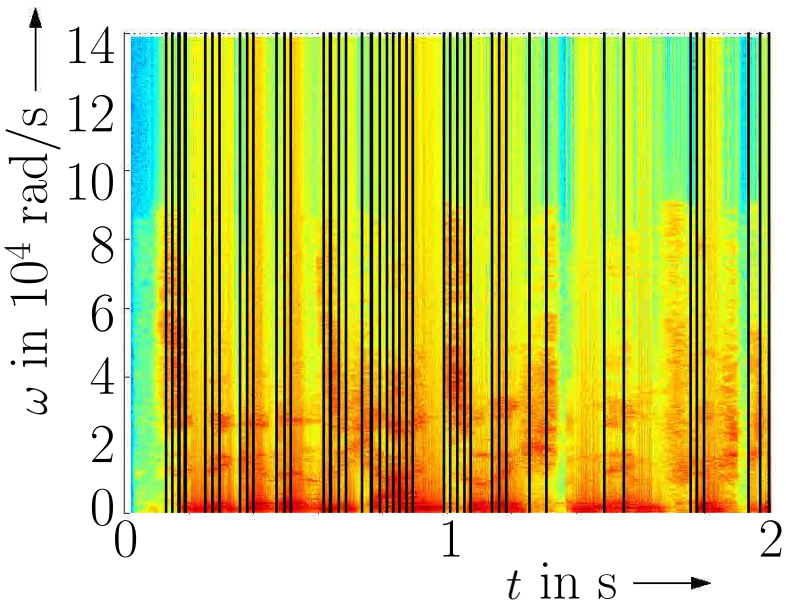
A. Rauh, S. Tiede, and C. Klenke. *Stochastic Filter Approaches for a Phoneme-Based Segmentation of Speech Signals.* In Proc. of 21st IEEE Intl. Conference on Methods and Models in Automation and Robotics MMAR 2016, Miedzyzdroje, Poland, 2016.

Phoneme-Based Segmentation of the Speech Signal (4)

Threshold test for $\Delta\tilde{\mu}^{e}_{x,k,3}$



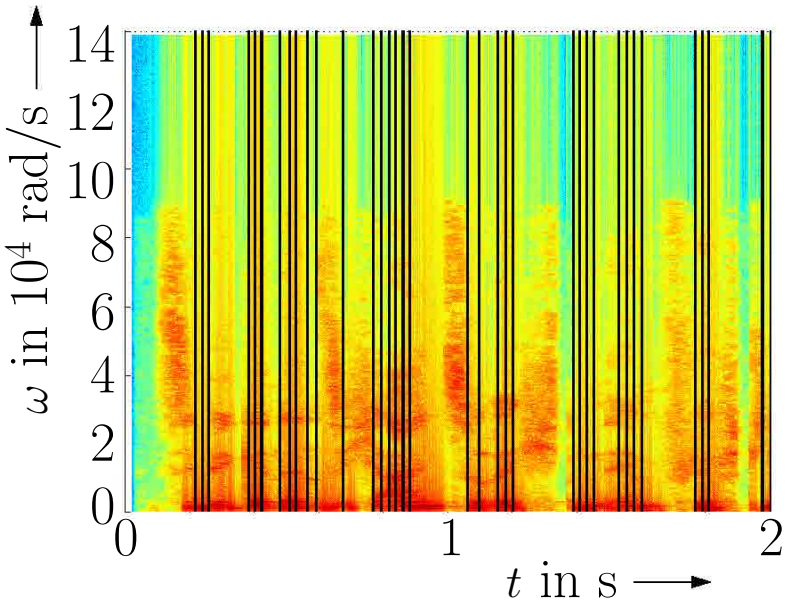
Threshold test for $\Delta\tilde{\mu}^{e}_{x,k,6}$



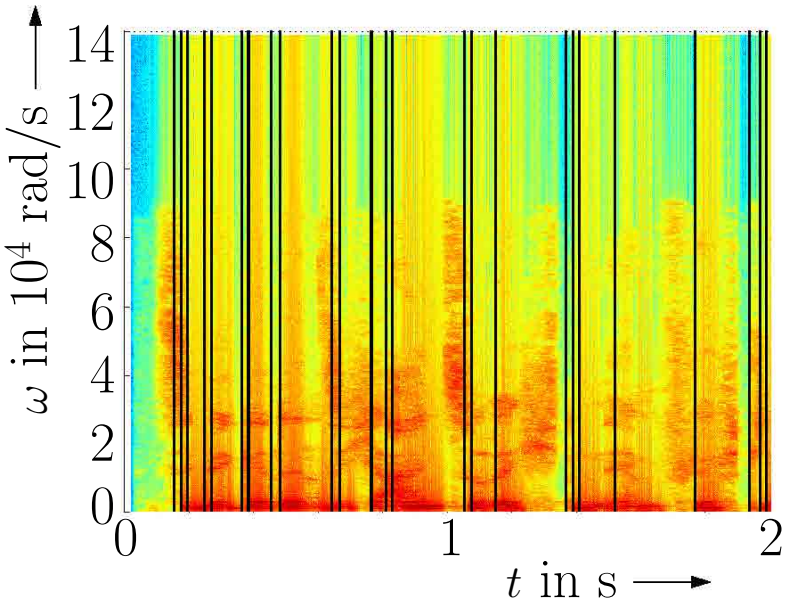
- Variations of the estimated frequencies
- Segmentation for the Extended Kalman Filter estimates with $n = 2$

Phoneme-Based Segmentation of the Speech Signal (5)

Threshold test for $\Delta\tilde{\mathbb{C}}^{e}_{x,k,(3,3)}$

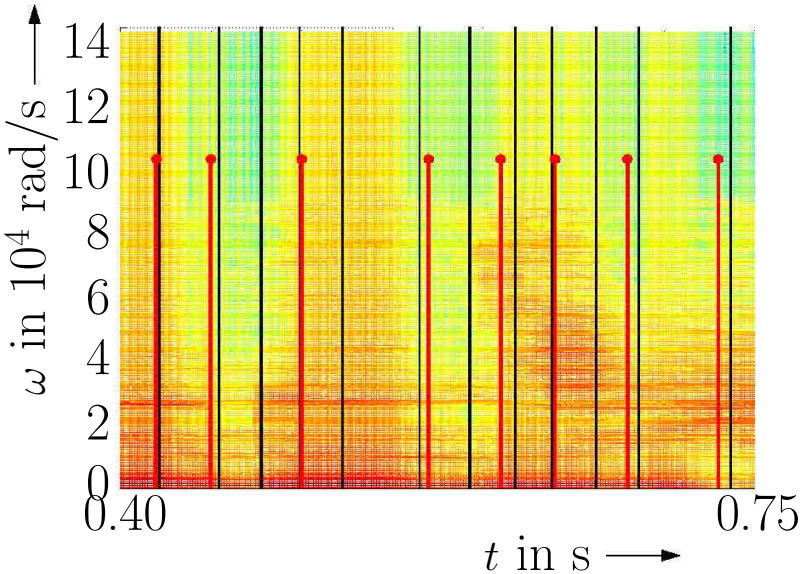


Threshold test for $\Delta\tilde{\mathbb{C}}^{e}_{x,k,(6,6)}$



- Variations of the estimated (co-)variances
- Segmentation for the Extended Kalman Filter estimates with $n = 2$

Phoneme-Based Segmentation of the Speech Signal (6)



- Merging of both criteria with a minimum temporal distance $\Delta T_{\min} = 20\,\mathrm{ms}$
- Comparison with a manually performed segmentation (red)

Note

- Accurate detection of phoneme boundaries
- Detection of characteristic variations within a single phoneme

Interval-Based Phoneme Database

Relevant Features

- Convex interval hull (CIH) over the expected values of each formant frequency
- CIH over the standard deviations of each formant frequency
- CIH over the amplitudes associated with each formant frequency
- Averaged intervals with respect to their corresponding duration (in case of multiple temporal subintervals per phoneme)
- Speaker-dependent normalization with respect to the mean of ω_1

Database of reference values

- Averaging of selected phoneme features from a reference data set (recording of TV news broadcast) for each of the phonemes

Interval-Based Phoneme Database

Relevant Features

- Convex interval hull (CIH) over the expected values of each formant frequency
- CIH over the standard deviations of each formant frequency
- CIH over the amplitudes associated with each formant frequency
- Averaged intervals with respect to their corresponding duration (in case of multiple temporal subintervals per phoneme)
- Speaker-dependent normalization with respect to the mean of ω_1

Database of reference values

- Future work: Test whether a subinterval representation (according to the automatic segmentation procedure) is more efficient.
- Then, a weighting factor can be assigned to each subinterval representing the relative number of occurrences as a further feature.

Threshold Distinction between Voiced and Unvoiced Sounds (1)

Unvoiced phonemes: Characterized by *large* estimated standard deviations (esp. for $i \geq 2$)

- Interval for normalized standard deviation of the second formant $[\tilde{\sigma}_2^e]$
- Duration of a phoneme τ
- Criterion 1:

$$\frac{\sup\{[\tilde{\sigma}_2^e]\} - \sigma^*}{\sigma^* - \inf\{[\tilde{\sigma}_2^e]\}} > 0.5$$

- Criterion 2:

$$(\inf\{[\tilde{\sigma}_2^e]\} > \sigma^*)\,\&\,(\tau < 65\,\mathrm{ms})$$

- Either Criterion 1 or Criterion 2 has to be fulfilled

Threshold Distinction between Voiced and Unvoiced Sounds (2)

Voiced phonemes: Characterized by *small* estimated standard deviations (esp. for $i \geq 2$)

- Interval for normalized standard deviation of the second formant $[\tilde{\sigma}_2^e]$
- Intervals for the estimated formant frequencies $[\tilde{\omega}_1^e]$, $[\tilde{\omega}_2^e]$
- Intervals for the estimated formant amplitudes $[\tilde{\alpha}_1^e]$, $[\tilde{\alpha}_2^e]$
- Duration of a phoneme τ
- Criterion:

$$(\sup\{[\tilde{\sigma}_2^e]\} < \sigma^*) \,\&\, (\inf\{[\tilde{\omega}_2^e]\} > \sup\{[\tilde{\omega}_1^e]\}) \ldots$$
$$\&\, (\tau \geq 50\,\mathrm{ms}) \,\&\, \left(\frac{\sup\{[\tilde{\alpha}_1^e]\}}{\sup\{[\tilde{\alpha}_2^e]\}} > 0.8\right)$$

Threshold Distinction between Voiced and Unvoiced Sounds (3)

Classification results for a test dataset: Phonemes classified as ...

- Unvoiced: `'S' 'n' 'n' 's' 'n' 'ch' 't'`
- Voiced: `'i' 'i' 'e' 'n'`
- Undecided: `'b' 'e' 'g' 'e' 'i' 'ch' 'ei' 'z' 'u' 'r' 'i'`

Note

- Misclassified `'n'` is hard to detect, because it follows a short `'i'`
- Vowels in the class *undecided* are reliably detected by the following phoneme classifier for vowels

Classification of Vowels

Similarity measure for vowels

- Short hand notation for an n-dimensional interval box:

$$\prod \left(\operatorname{diam}\{[\mathbf{x}]\}\right) = (\overline{x}_1 - \underline{x}_1) \cdot \ldots \cdot (\overline{x}_n - \underline{x}_n)$$

- Current feature box $[\mathbf{x}]$
- Reference feature box $[\mathbf{x}]_{\text{ref}}$
- Convex interval hull denoted by $\cup$
- Similarity measure

$$\frac{\prod \left(\operatorname{diam}\{[\mathbf{x}]\}\right) + \prod \left(\operatorname{diam}\{[\mathbf{x}]_{\text{ref}}\}\right)}{\prod \left(\operatorname{diam}\{[\mathbf{x}] \cup [\mathbf{x}]_{\text{ref}}\}\right)}$$

Classification of Vowels — Results

Table of similarity measures

	'i'	'e'	'i'	'e'	'i'	'u'	'i'	'e'
'e'	0.3823	0.0220	0.0163	0.0225	0.0526	0.1679	0.0022	0.0246
'i'	0.1527	0.0092	0.1952	0.4986	0.1008	0.0572	0.0176	0.0382
'u'	0.1129	0.0586	0.0344	0.0563	0.0233	2.0000	0.0096	0.0756

Classification of Vowels — Results

Table of similarity measures

	'i'	'e'	'i'	'e'	'i'	'u'	'i'	'e'
'e'	0.3823	0.0220	0.0163	0.0225	0.0526	0.1679	0.0022	0.0246
'i'	0.1527	0.0092	0.1952	0.4986	0.1008	0.0572	0.0176	0.0382
'u'	0.1129	0.0586	0.0344	0.0563	0.0233	2.0000	0.0096	0.0756

Relative similarity measures in percent

	'i'	'e'	'i'	'e'	'i'	'u'	'i'	'e'
'e'	59.002	24.479	6.626	3.892	29.763	7.546	7.345	17.804
'i'	23.568	10.203	79.383	86.350	57.042	2.572	59.909	27.577
'u'	17.430	65.318	13.991	9.758	13.195	89.882	32.746	54.619

Classification of Vowels — Results

Table of similarity measures

	'i'	'e'	'i'	'e'	'i'	'u'	'i'	'e'
'e'	0.3823	0.0220	0.0163	0.0225	0.0526	0.1679	0.0022	0.0246
'i'	0.1527	0.0092	0.1952	0.4986	0.1008	0.0572	0.0176	0.0382
'u'	0.1129	0.0586	0.0344	0.0563	0.0233	2.0000	0.0096	0.0756

Relative similarity measures in percent

	'i'	'e'	'i'	'e'	'i'	'u'	'i'	'e'
'e'	∗59.002	24.479	6.626	3.892	29.763	7.546	7.345	17.804
'i'	23.568	10.203	79.383	∗86.350	57.042	2.572	59.909	27.577
'u'	17.430	∗65.318	13.991	9.758	13.195	89.882	32.746	∗54.619

Remaining misclassifications will be removed in future work by subinterval representations of phonemes $\Longrightarrow$ Especially for bilabial stops ('b','p')

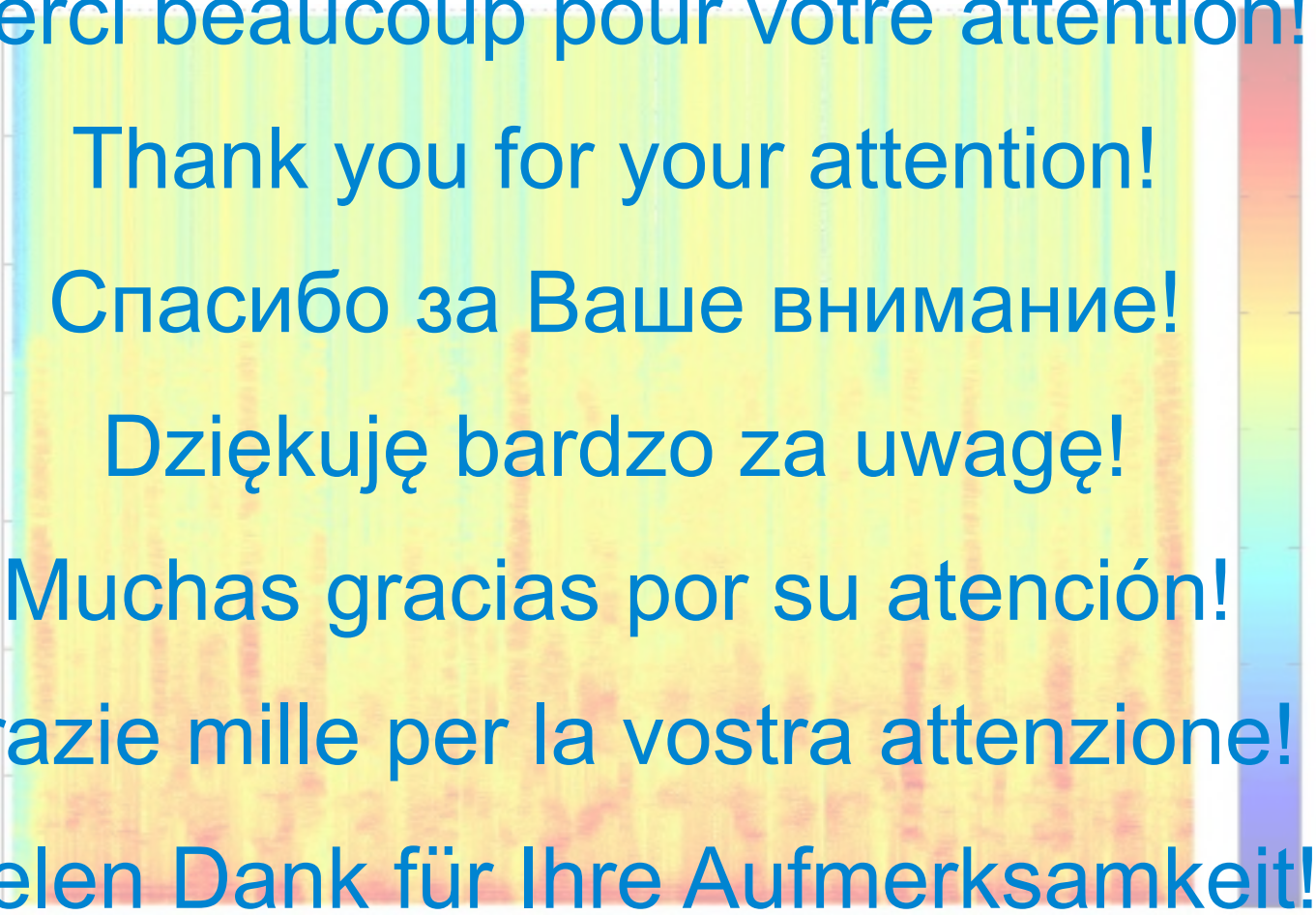
Conclusions and Outlook on Future Work

- Stochastic filtering approach for the online frequency estimation of speech signals
- Stochastic filter as the basis for the phoneme-based segmentation of speech signals
- Fundamental interval representation of phoneme features
- Basic classifier distinguishing between unvoiced and voiced sounds

Conclusions and Outlook on Future Work

- Stochastic filtering approach for the online frequency estimation of speech signals
- Stochastic filter as the basis for the phoneme-based segmentation of speech signals
- Fundamental interval representation of phoneme features
- Basic classifier distinguishing between unvoiced and voiced sounds

- Extension of the classification procedure
  - Further interval-based (pre-)classification stages
  - Stochastic classification by multi-hypothesis Kalman Filters
- Validation on further test datasets without and with pronunciation disorders



Merci beaucoup pour votre attention!

Thank you for your attention!

Спасибо за Ваше внимание!

Dziękuję bardzo za uwagę!

¡Muchas gracias por su atención!

Grazie mille per la vostra attenzione!

Vielen Dank für Ihre Aufmerksamkeit!