Automatic Pathological Speech Intelligibility Assessment Exploiting Subspace-Based Analyses

Parvaneh Janbakhshi, Ina Kodrasi, and Hervé Bourlard

Idiap Research Institute
Speech and Audio Processing Group

June 2020

Outline

1. Automatic Pathological speech intelligibility assessment

2. Subspace-based pathological speech intelligibility assessment
   - Proposed SBI measure
   - Modulation spectrum and speech intelligibility
   - Empirical insights into SBI measure

3. Experimental results

4. Summary

Automatic pathological speech intelligibility assessment (State-of-the-art)

- Our previously propsoed intelligibility measures based on short-time objective measure (P-ESTOI) (Janbakhshi et al., 2019a)
  - Simple structure
  - No need for a large amount of healthy speech data
  - Based on a single feature without training (no risk of overfitting)
  - Generalizable across languages and neurological diseases

Automatic pathological speech intelligibility assessment (State-of-the-art)

- Our previously propsoed intelligibility measures based on short-time objective measure (P-ESTOI) (Janbakhshi et al., 2019a)
  - Simple structure
  - No need for a large amount of healthy speech data
  - Based on a single feature without training (no risk of overfitting)
  - Generalizable across languages and neurological diseases

- P-ESTOI drawbacks
  - Requires time-alignment by dynamic time warping (DTW)
    - Failure of time-alignment for severe patients
    - Requires healthy recordings matching the phonetic content of the pathological speech $\Rightarrow$ Not applicable to phonetically unbalanced scenarios

Objectives

- Goal ⇒ Automatic intelligibility assessment of pathological speech
  - Reliable and efficient measures
  - Without requiring time-alignment and large amounts of training data
  - Applicable to phonetically unbalanced scenarios

Objectives

- Goal ⇒ Automatic intelligibility assessment of pathological speech
  - Reliable and efficient measures
  - Without requiring time-alignment and large amounts of training data
  - Applicable to phonetically unbalanced scenarios
- Proposing intelligibility measures motivated by P-ESTOI while avoiding its drawbacks:
  - Spectral subspace analysis of speech (Janbakhshi et al., 2020, 2019b)

Subspace-based intelligibility (SBI) measure

Extension of our previous work (Janbakhshi et al., 2019b)

Subspace-based intelligibility (SBI) measure

Extension of our previous work (Janbakhshi et al., 2019b)

- Characterizing healthy and pathological speech signals by spectral subspaces of (possibly) different dimensions

Subspace-based intelligibility (SBI) measure

Extension of our previous work (Janbakhshi et al., 2019b)

- Characterizing healthy and pathological speech signals by spectral subspaces of (possibly) different dimensions
- Providing empirical evidence on:
  - Relation between the SBI measure and low-frequency components of the spectral modulation of speech
  - Robustness of the SBI measure to gender variations
  - Robustness of the SBI measure to age variations

Subspace-based intelligibility (SBI) measure

Extension of our previous work (Janbakhshi et al., 2019b)

- Characterizing healthy and pathological speech signals by spectral subspaces of (possibly) different dimensions
- Providing empirical evidence on:
  - Relation between the SBI measure and low-frequency components of the spectral modulation of speech
  - Robustness of the SBI measure to gender variations
  - Robustness of the SBI measure to age variations
- Incorporating short-time temporal information in the SBI measure

Subspace-based intelligibility (SBI) measure

Extension of our previous work (Janbakhshi et al., 2019b)

- Characterizing healthy and pathological speech signals by spectral subspaces of (possibly) different dimensions
- Providing empirical evidence on:
  - Relation between the SBI measure and low-frequency components of the spectral modulation of speech
  - Robustness of the SBI measure to gender variations
  - Robustness of the SBI measure to age variations
- Incorporating short-time temporal information in the SBI measure
- Experimental evaluation of intelligibility assessment:
  - Two scenarios: phonetically balanced and unbalanced scenarios
  - Two languages, i.e., English and Dutch, and two pathologies, i.e.,CP and HI

Outline

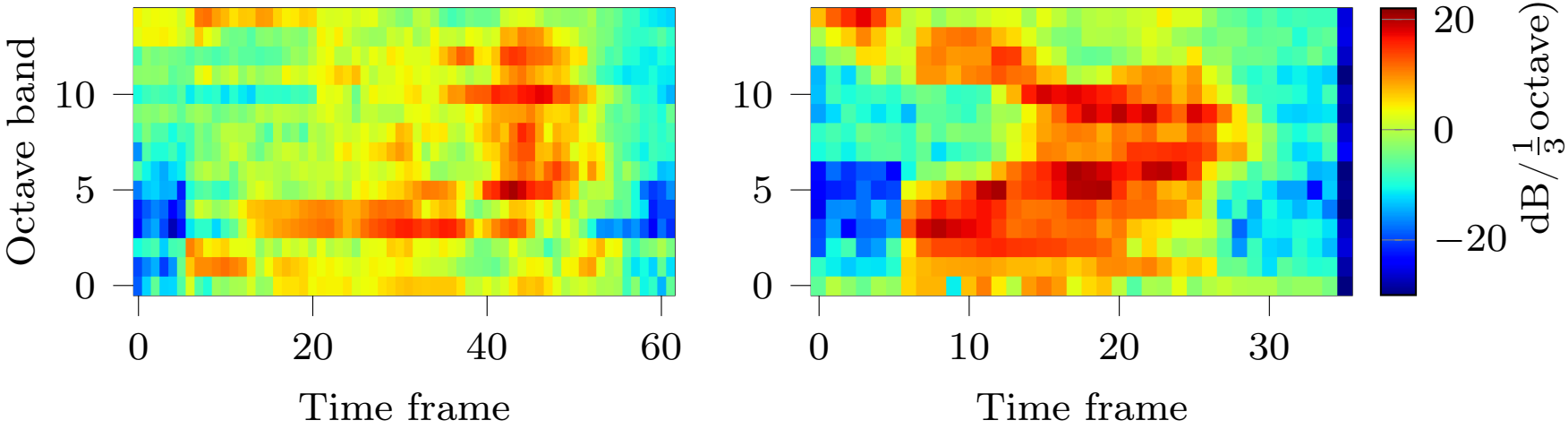
1. Automatic Pathological speech intelligibility assessment

2. Subspace-based pathological speech intelligibility assessment
   - Proposed SBI measure
   - Modulation spectrum and speech intelligibility
   - Empirical insights into SBI measure

3. Experimental results

4. Summary

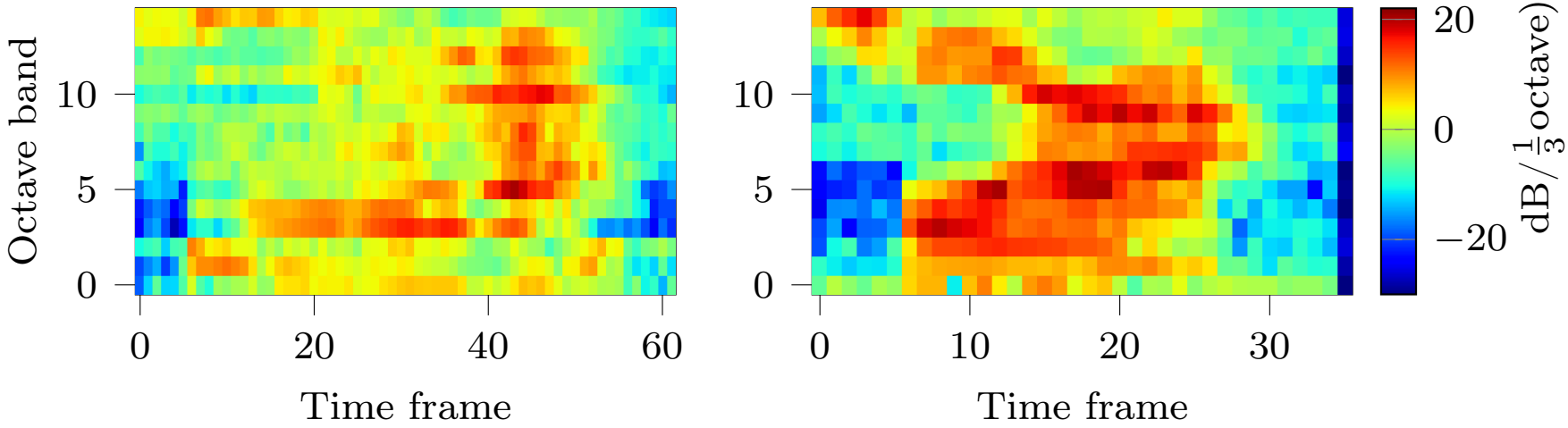
One-third octave band representation of speech



(a) Low-intelligible patient

(b) Healthy speaker

One-third octave band representation of speech



(a) Low-intelligible patient

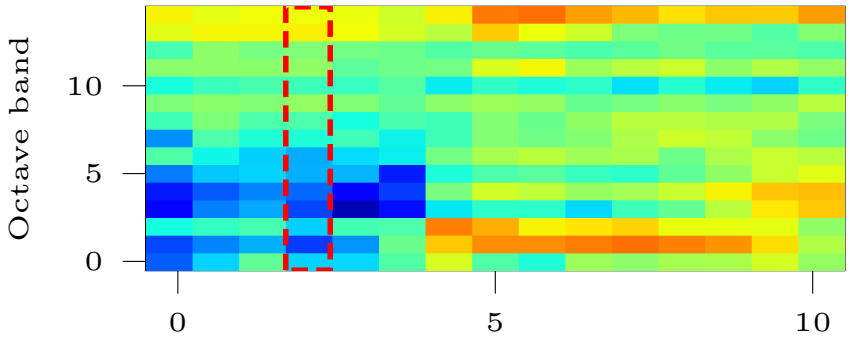
(b) Healthy speaker

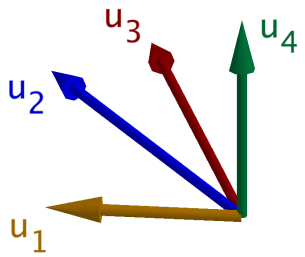
Main hypotheses

✓ Dominant spectral patterns of intelligible (healthy) speech differs from pathological speech

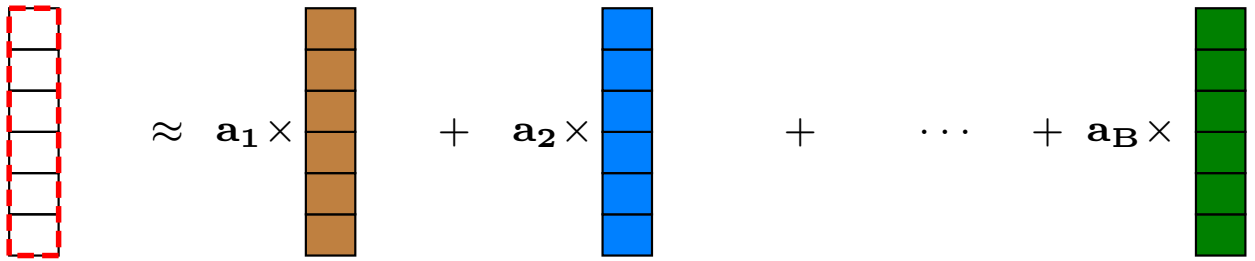
✓ Difference increases as pathological speech intelligibility decreases ⇒ Difference degree ∝ Intelligbility

Spectral basis vectors



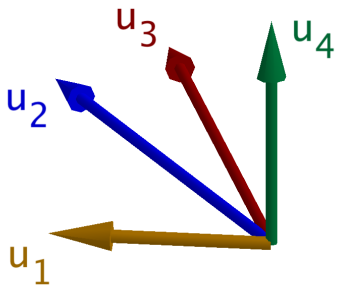


Spectral basis vectros of speech

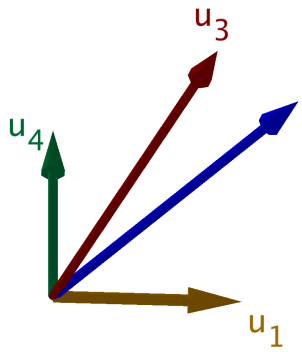


Spectral vector (at time frame i = 3) **Spectral basis 1 $\mathbf{u_1}$** **Spectral basis 2 $\mathbf{u_2}$** **Spectral basis B $\mathbf{u_B}$**

Spectral basis vectors

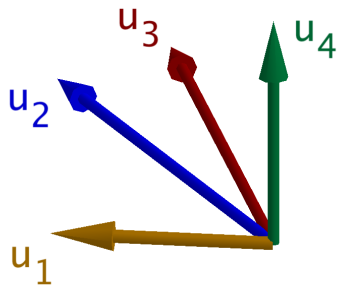


Spectral basis vectors from healthy speech (intelligible reference)

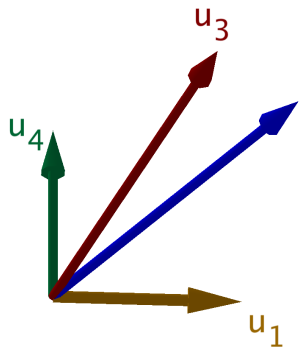


Spectral basis vectors from pathological speech (pathological test)

Spectral basis vectors



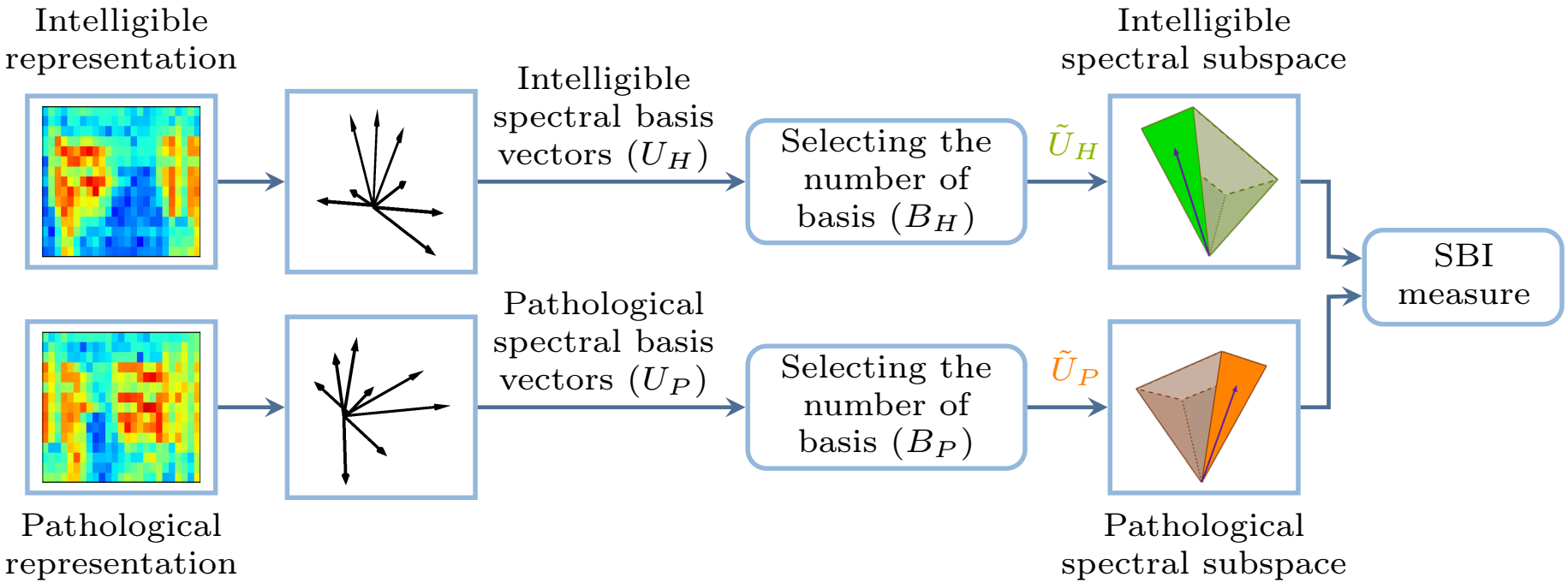
Spectral basis vectors from healthy speech (intelligible reference)



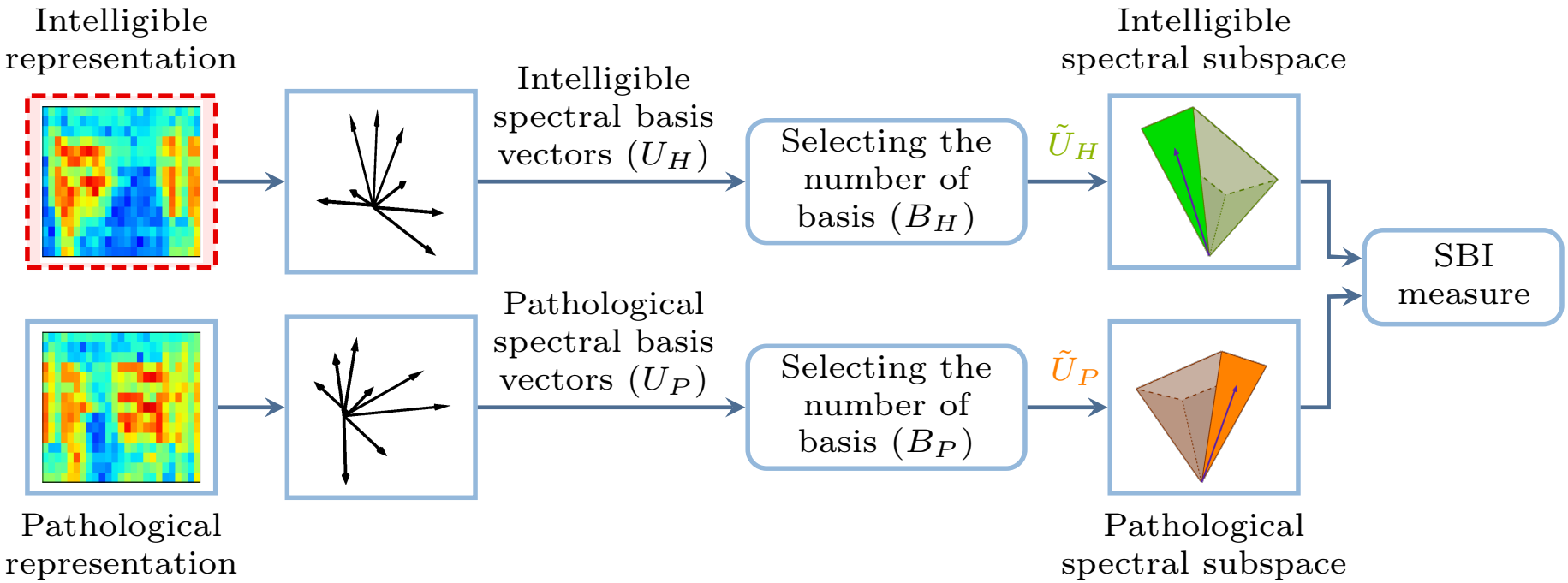
Spectral basis vectors from pathological speech (pathological test)

- How to find spectral basis vectors?
- How many spectral basis vectors?
- How to quantify the distance (difference) between pathological and healthy spectral basis vectors?

SBI measure


Intelligible representation
Intelligible spectral basis vectors (U_H)
Selecting the number of basis (B_H)
$\tilde{U}_H$
Intelligible spectral subspace
SBI measure
Pathological spectral basis vectors (U_P)
Selecting the number of basis (B_P)
$\tilde{U}_P$
Pathological representation
Pathological spectral subspace

SBI measure

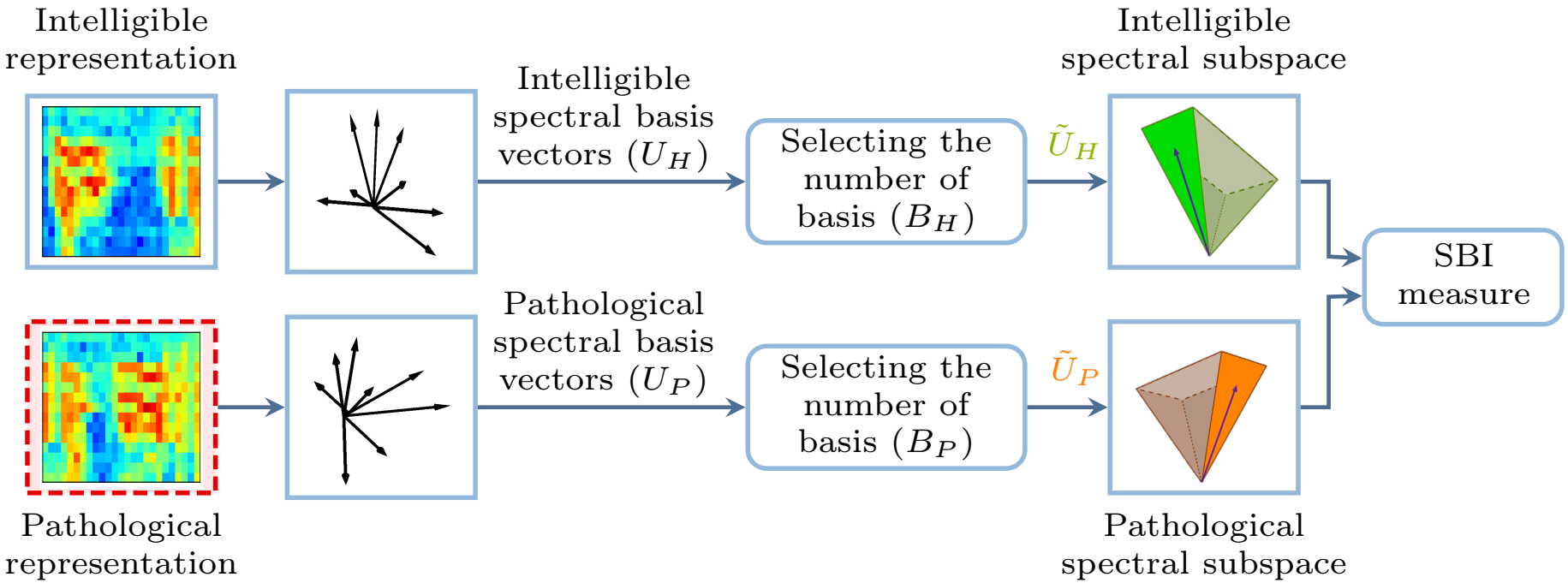


TF domain: logarithm of the one-third octave band spectrum

- Healthy reference representations $\mathbf{H}$ from:
  - *possibly but not necessarily the same utterances* from S different healthy speakers:

$\mathbf{H} = [\mathbf{H}_1\ \mathbf{H}_2\ \ldots\ \mathbf{H}_S]$, with $\mathbf{H}_s$ from healthy speaker s

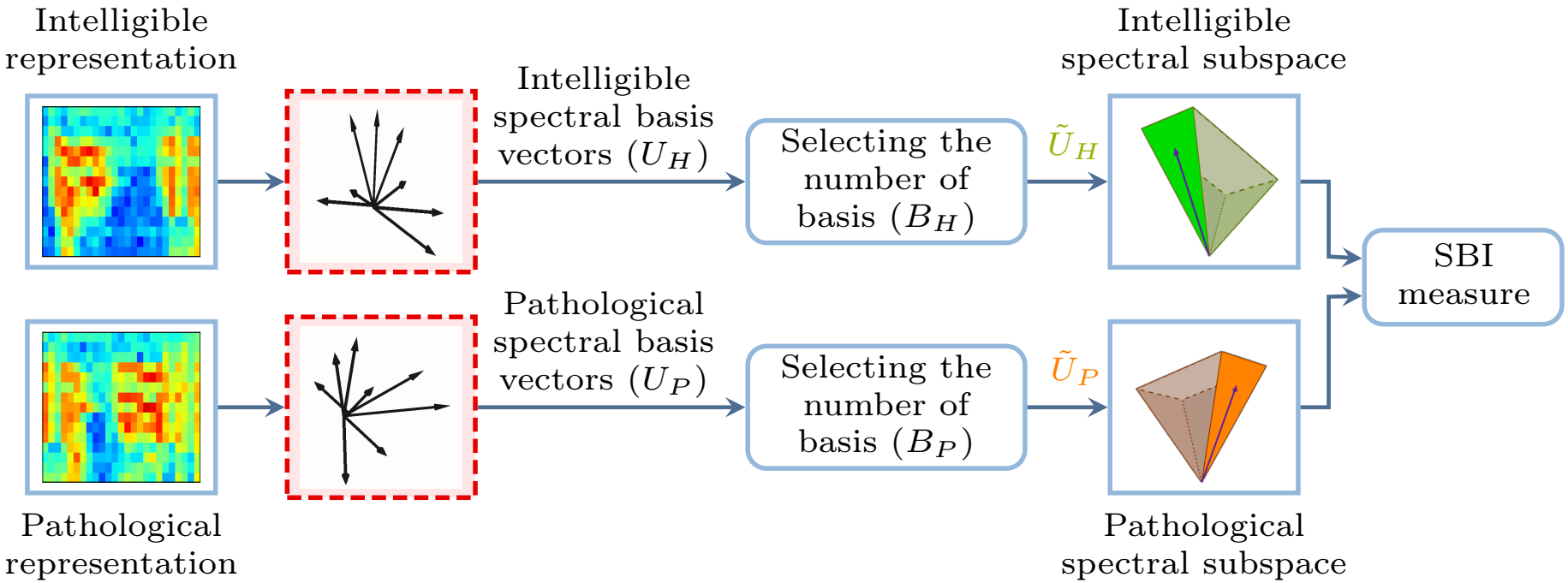
SBI measure



TF domain: logarithm of the one-third octave band spectrum

- Test (i.e., pathological) representations $\mathbf{P}_r$ from
  - test utterance from patient r *(possibly but not necessarily the same utterance as healthy speakers)*

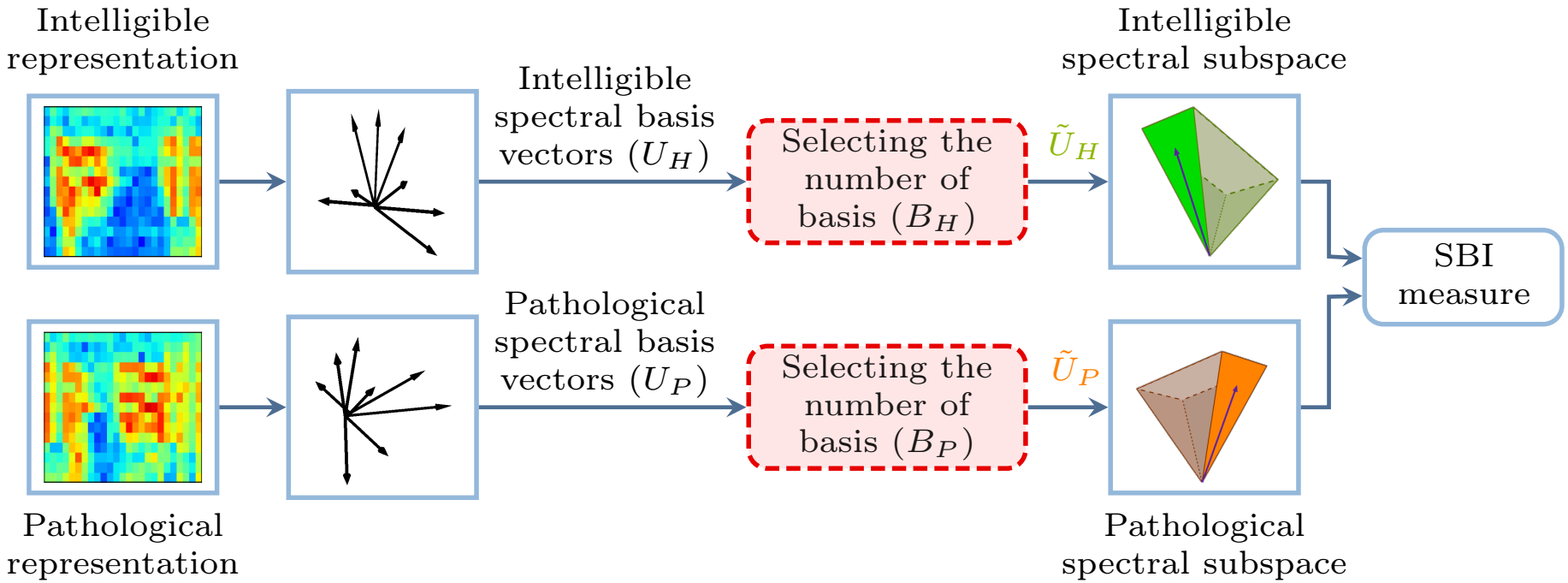
SBI measure



Computing spectral basis vectors by singular value decomposition (SVD)

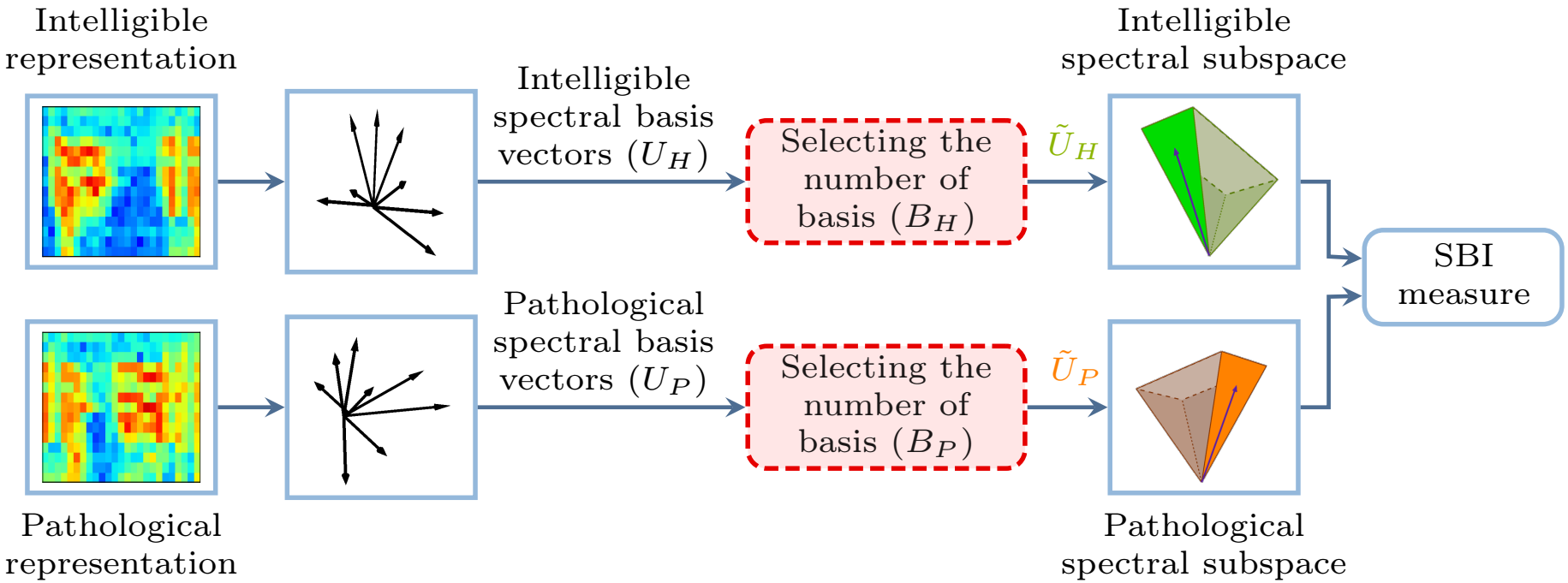
- SVD of healthy reference representations $\Rightarrow \mathbf{H} = \mathbf{U}_H \mathbf{\Sigma}_H \mathbf{V}_H^T$
- SVD of test (i.e., pathological) representations $\Rightarrow \mathbf{P}_r = \mathbf{U}_P \mathbf{\Sigma}_P \mathbf{V}_P^T$

SBI measure

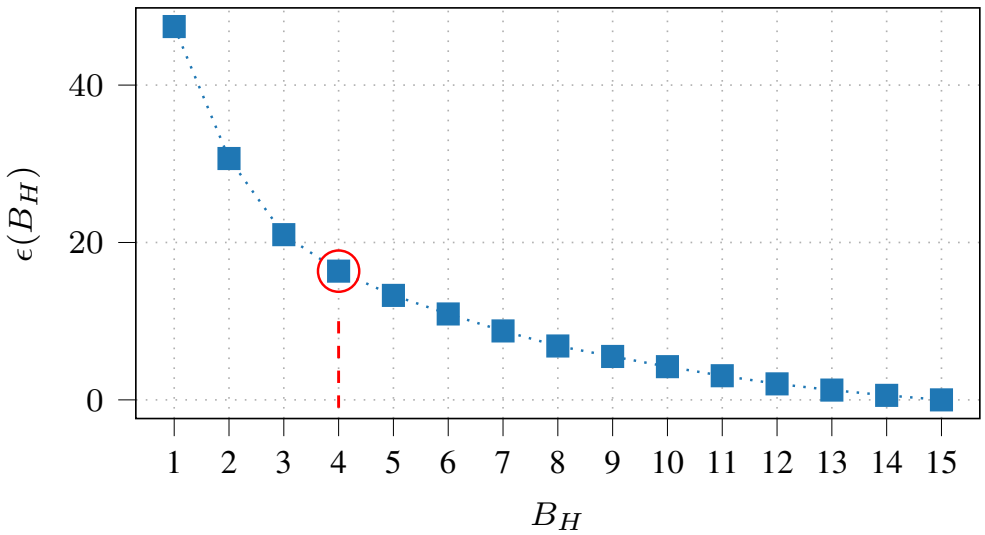


- Intelligible spectral basis:
  - First B_H (with $B_H < J$) dominant spectral basis vectors in $\mathbf{U}_H \Rightarrow \tilde{\mathbf{U}}_H$
- Test (i.e., pathological) spectral basis:
  - First B_P (with $B_P < J$) dominant spectral basis vectors in $\mathbf{U}_P \Rightarrow \tilde{\mathbf{U}}_P$

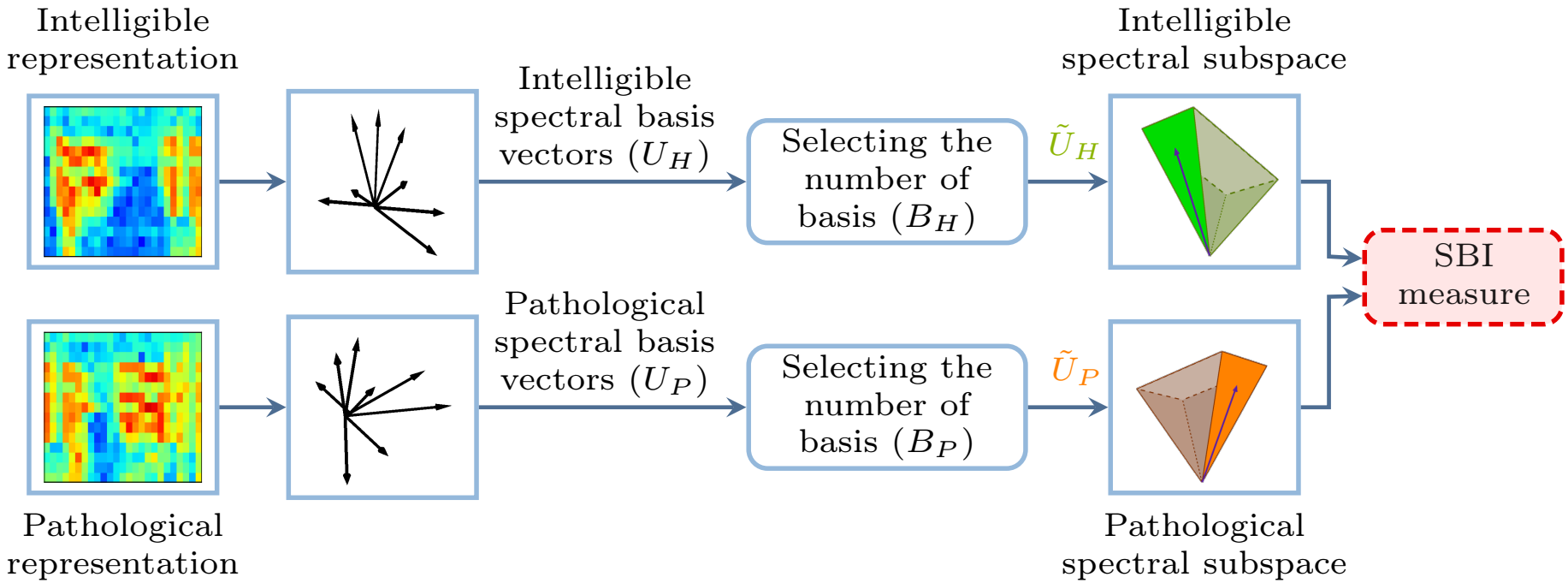
SBI measure



- Automatic selection of B_H & B_P
  - L-curve method
  - Approximation error plot $\epsilon(B_H)$ & $\epsilon(B_P)$
  - Possibly $B_H \neq B_P$



SBI measure



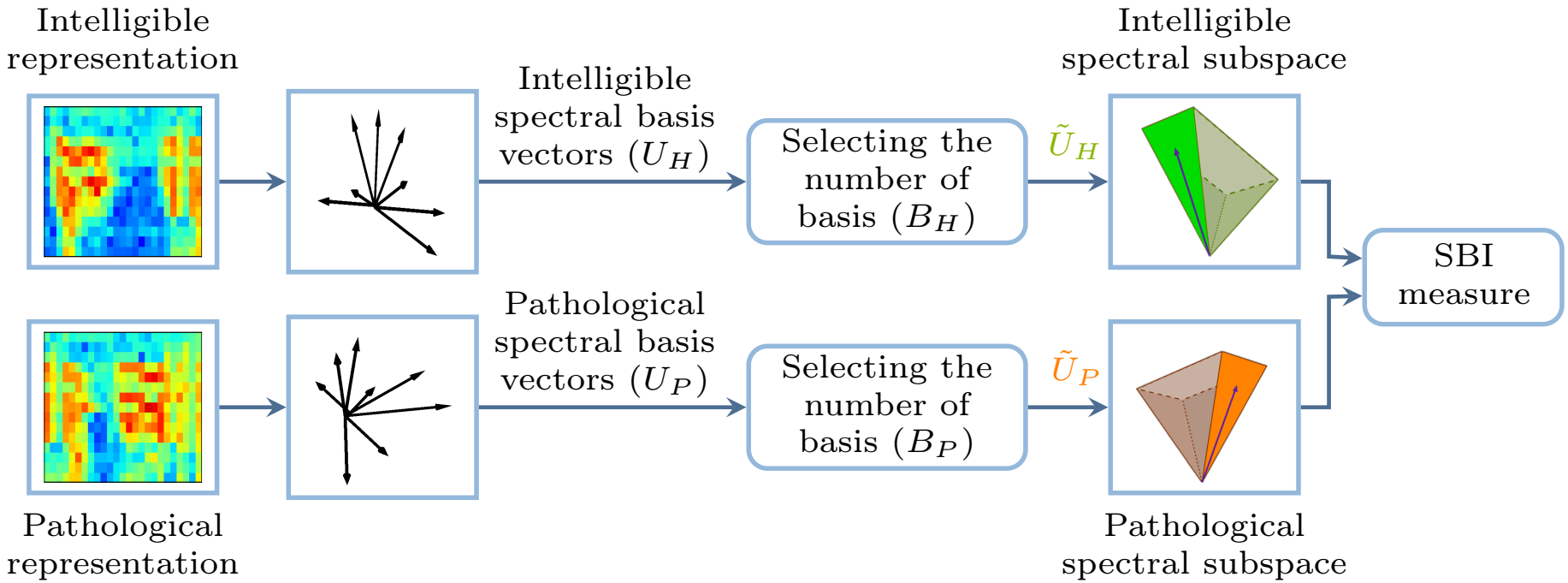
» Quantifying the distance between (1) and (2) ⇒ SBI score:

- (1) the subspaces spanned by $\tilde{\mathbf{U}}_H$ (intelligible spectral subspace)
- (2) the subspaces spanned by $\tilde{\mathbf{U}}_P$ (pathological spectral subspace)

$$\delta(\tilde{\mathbf{U}}_H, \tilde{\mathbf{U}}_P) = 2\sqrt{\sum_{i=1}^{\min(B_H, B_P)} sin^2(\theta_i/2)},$$

(θ_i: the i^{th} principal angle between subspaces)

Incorporating temporal information in SBI



SBI ignores temporal patterns

Incorporate temporal information for intelligibility assessment?

» Long-time temporal patterns ⇒ Temporal basis ? ✗

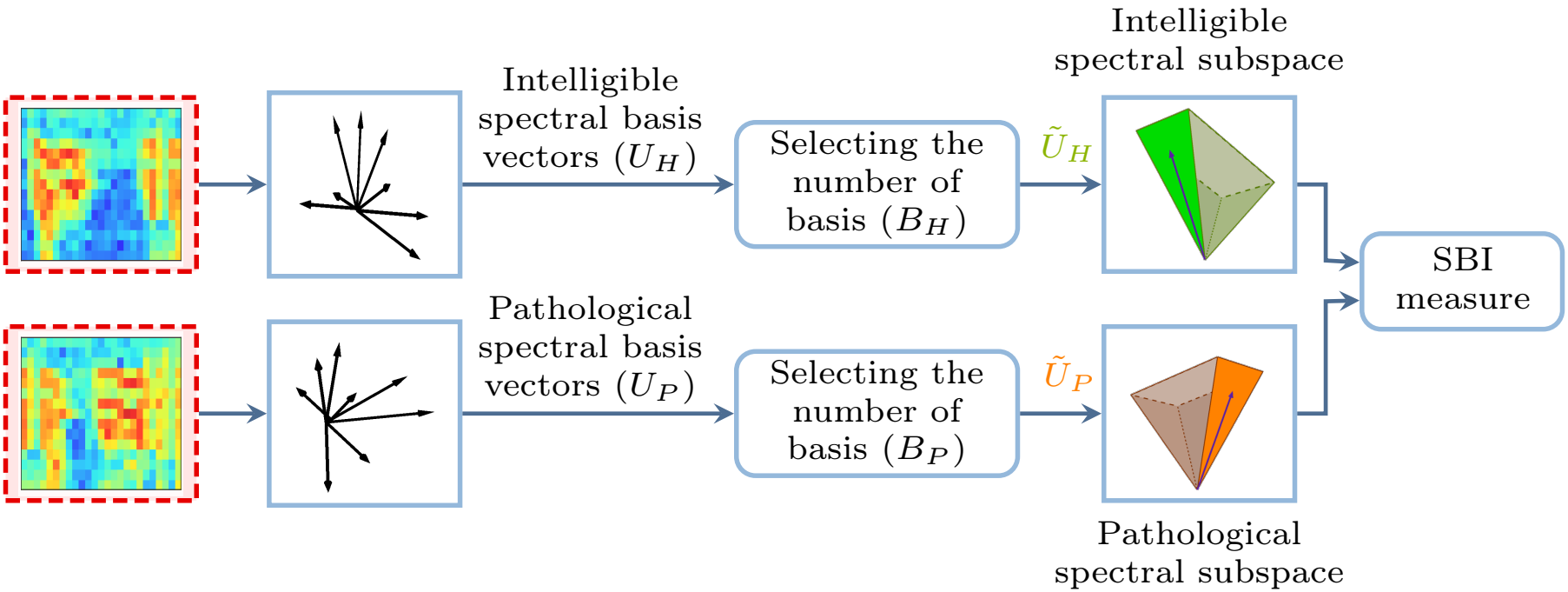
- Not possible because of unaligned representations

SVD
$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$

» Short-time temporal patterns ? ✓

- Modifying the TF representations to consider short-time temporal patterns

Incorporating temporal information in SBI



» Dynamic SBI measure

$$\mathbf{H}_{\text{DSBI}} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_{d+1} & \dots & \mathbf{h}_{(k-1)d+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{h}_d & \mathbf{h}_{2d} & \dots & \mathbf{h}_{kd} \end{bmatrix}$$

» Moving average SBI measure

$$\mathbf{H}_{\text{MASBI}} = [\mathbf{h}'_1 \quad \mathbf{h}'_2 \quad \dots \quad \mathbf{h}'_{M-q+1}],$$

$$\mathbf{h}'_m = \frac{1}{q} \sum_{j=m}^{m+q-1} \mathbf{h}_j$$

Modulation spectrum and speech intelligibility

- Fluctuations of the speech power spectrogram in time (at any frequency) $\Rightarrow$ temporal modulations
- Fluctuations of the speech power spectrogram in frequency (at any time frame) $\Rightarrow$ spectral modulations

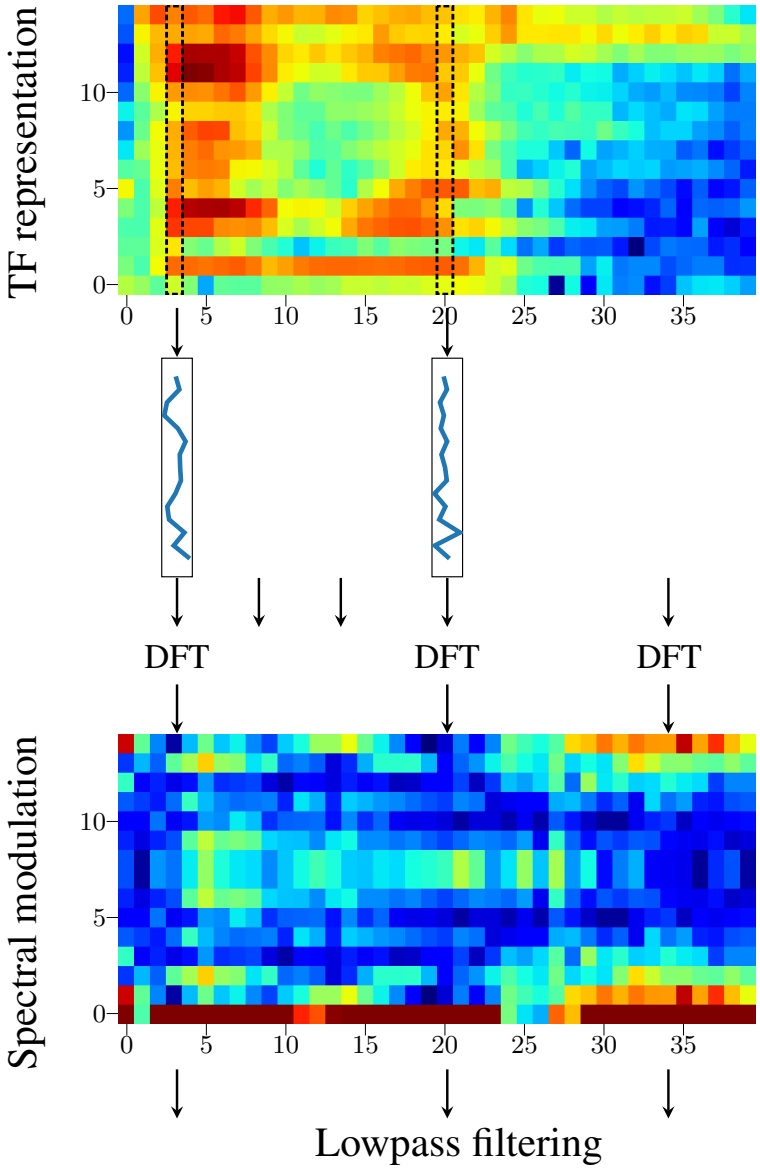
Modulation spectrum and speech intelligibility

- Fluctuations of the speech power spectrogram in time (at any frequency) $\Rightarrow$ temporal modulations
- Fluctuations of the speech power spectrogram in frequency (at any time frame) $\Rightarrow$ spectral modulations
- Spectro-temporal modulations of speech are critical to speech perception
  - According to psychoacoustic studies
  - Success of many objective intelligibility measures based on modulation cues

Modulation spectrum and speech intelligibility

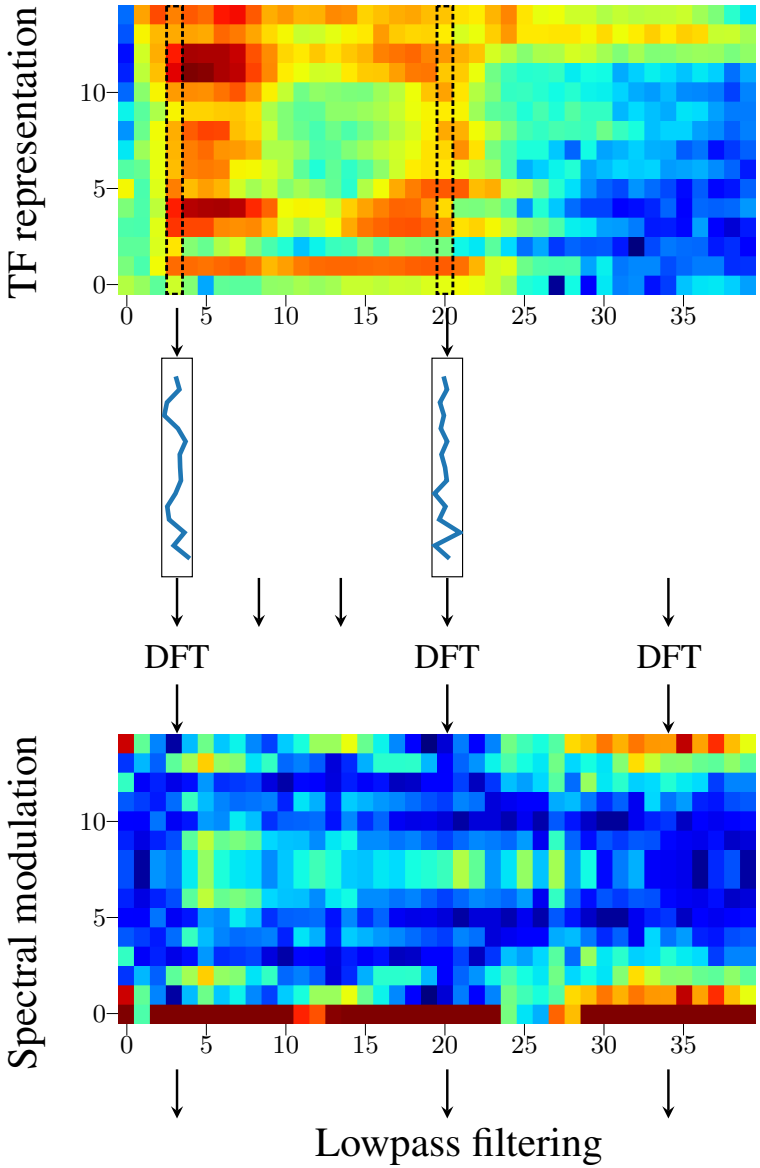
- Fluctuations of the speech power spectrogram in time (at any frequency) $\Rightarrow$ temporal modulations
- Fluctuations of the speech power spectrogram in frequency (at any time frame) $\Rightarrow$ spectral modulations
- Spectro-temporal modulations of speech are critical to speech perception
  - According to psychoacoustic studies
  - Success of many objective intelligibility measures based on modulation cues
- Can SBI measure reflect **important spectral modulation** differences between healthy and pathological speech?

Modulation spectrum and speech intelligibility (psychoacoustic study) (Elliott and Theunissen, 2009)



Effect of spectral modulation on subjective speech intelligibility

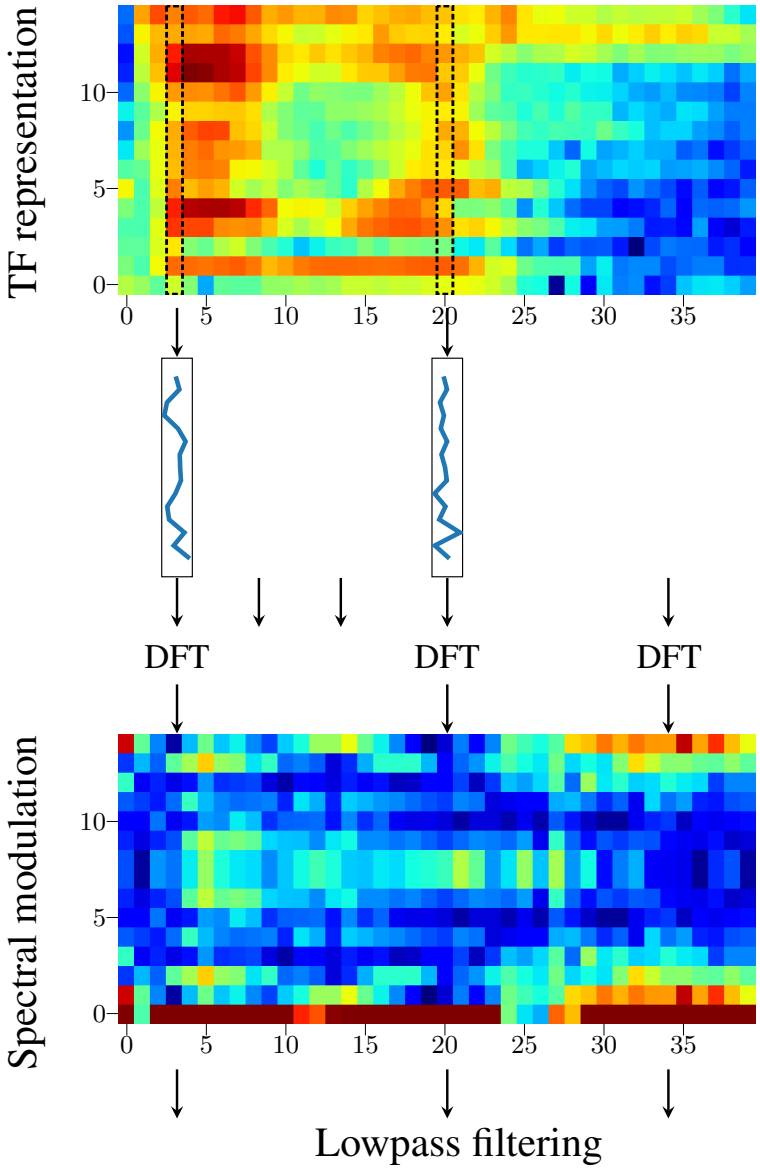
Modulation spectrum and speech intelligibility (psychoacoustic study) (Elliott and Theunissen, 2009)



Effect of spectral modulation on subjective speech intelligibility

- Fourier transform of each time frame
  $\Rightarrow$ spectral modulation pattern
  (in units of cycle/kHz, cycle/$\frac{1}{3}$octave)

Modulation spectrum and speech intelligibility (psychoacoustic study) (Elliott and Theunissen, 2009)

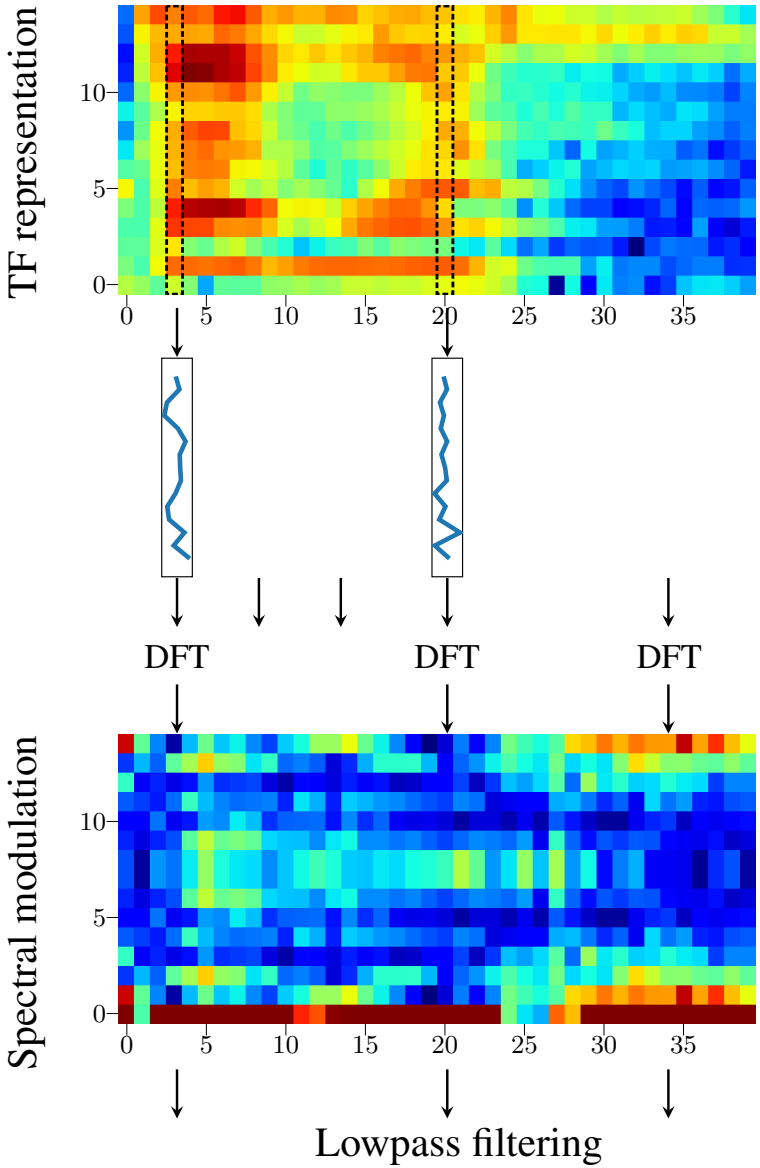


Effect of spectral modulation on subjective speech intelligibility

- Fourier transform of each time frame
  $\Rightarrow$ spectral modulation pattern
  (in units of cycle/kHz, cycle/$\frac{1}{3}$octave)

- Low-pass filtering the modulation spectrum at each time frame at different cut-off frequencies

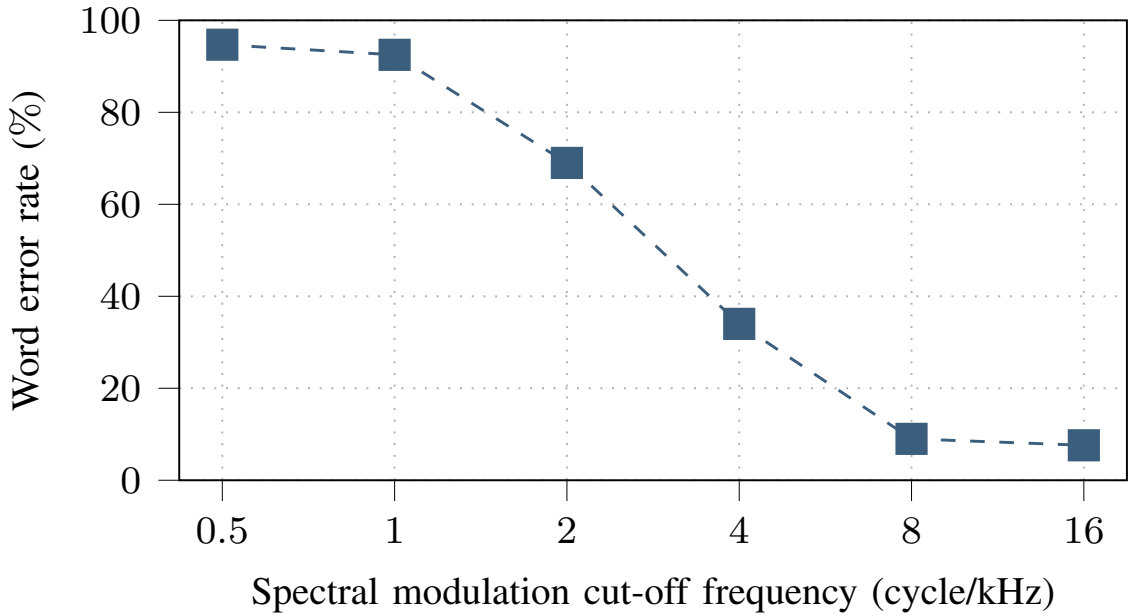
Modulation spectrum and speech intelligibility (psychoacoustic study) (Elliott and Theunissen, 2009)



Effect of spectral modulation on subjective speech intelligibility

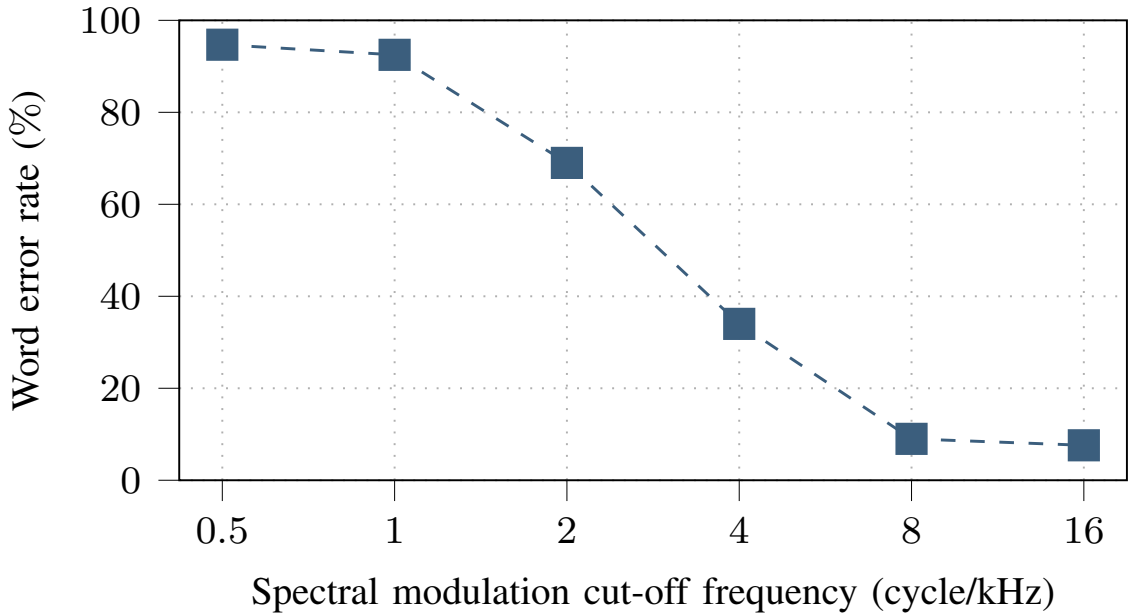
- Fourier transform of each time frame $\Rightarrow$ spectral modulation pattern (in units of cycle/kHz, cycle/$\frac{1}{3}$octave)
- Low-pass filtering the modulation spectrum at each time frame at different cut-off frequencies
- Rating intelligibility of synthetically manipulated utterances by human listeners

Modulation spectrum and speech intelligibility (psychoacoustic study) (Elliott and Theunissen, 2009)



- Removing low spectral modulation frequencies ⇒ significantly increasing word error rate, i.e., decreasing speech intelligibility
- Low-frequency spectral modulations contribute (the most) to the perceived speech intelligibility by human listeners

Modulation spectrum and speech intelligibility (psychoacoustic study) (Elliott and Theunissen, 2009)



- Removing low spectral modulation frequencies ⇒ significantly increasing word error rate, i.e., decreasing speech intelligibility
- Low-frequency spectral modulations contribute (the most) to the perceived speech intelligibility by human listeners
- Can SBI measure respond to missing spectral modulation frequencies similarly to humans' perception ???

Empirical insights into SBI measure

Empirical analyses of SBI

- SBI and spectral modulation of speech
- Robustness of SBI to gender and age variations

Used database

- PC-GITA database (Orozco et al., 2014)
  - 50 Spanish-speaking healthy speakers (25 males and 25 females) with age of the speakers ranging from 31 to 86 years old (median: 62)
  - 10 sentences from each speaker

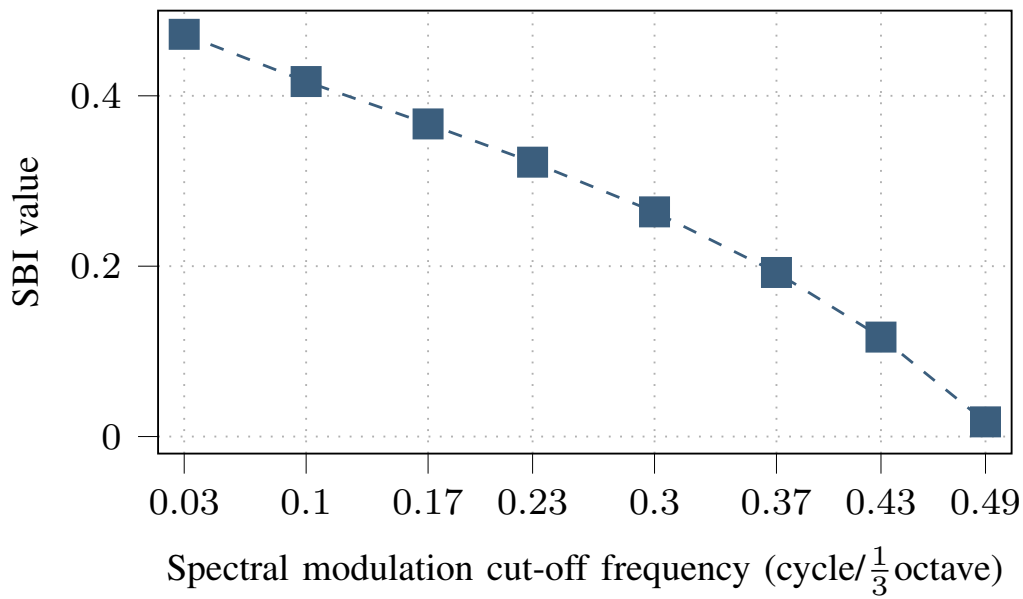
SBI and spectral modulation of speech

Analyzing the effect of spectral modulation cues on SBI measure (similar to the psychoacoustic study)

- Low-pass filtering spectral modulations at different cut-off frequencies
- Computing SBI measure based on distance between subspaces (1) and (2)
  - (1) Spectral subspaces spanning the original utterances
  - (2) Spectral subspaces spanning the low-pass spectral modulation filtered utterances

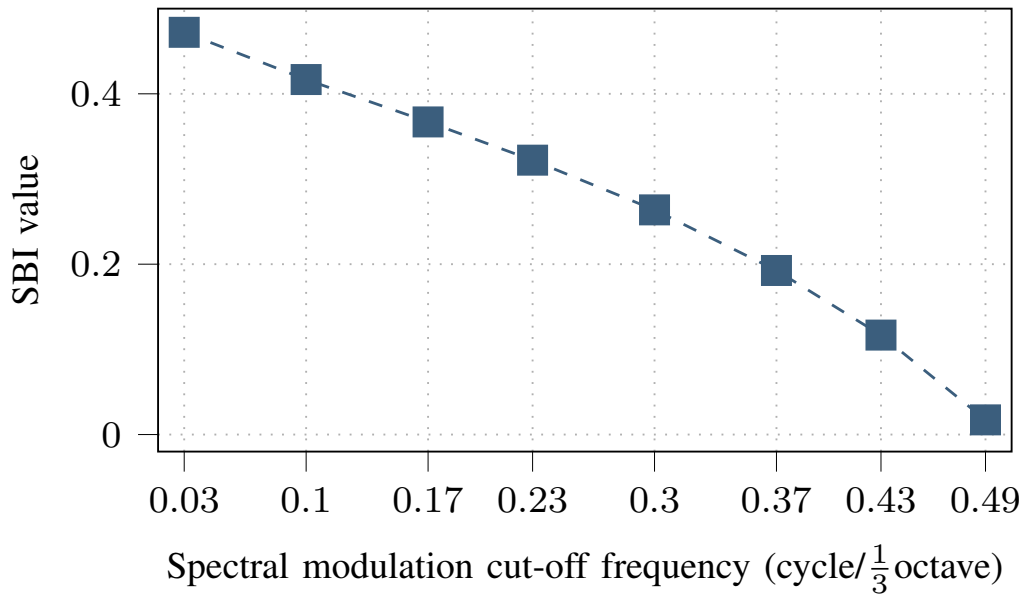
SBI and spectral modulation of speech

Analyzing the effect of spectral modulation cues on SBI measure (similar to the psychoacoustic study)



SBI and spectral modulation of speech

Analyzing the effect of spectral modulation cues on SBI measure (similar to the psychoacoustic study)



- The effect of missing spectral modulation frequencies **on SBI and on subjective intelligibility is similar**
- Low-frequency components of spectral modulations are **crucial** for speech intelligibility assessment through **both SBI and subjective intelligibility**

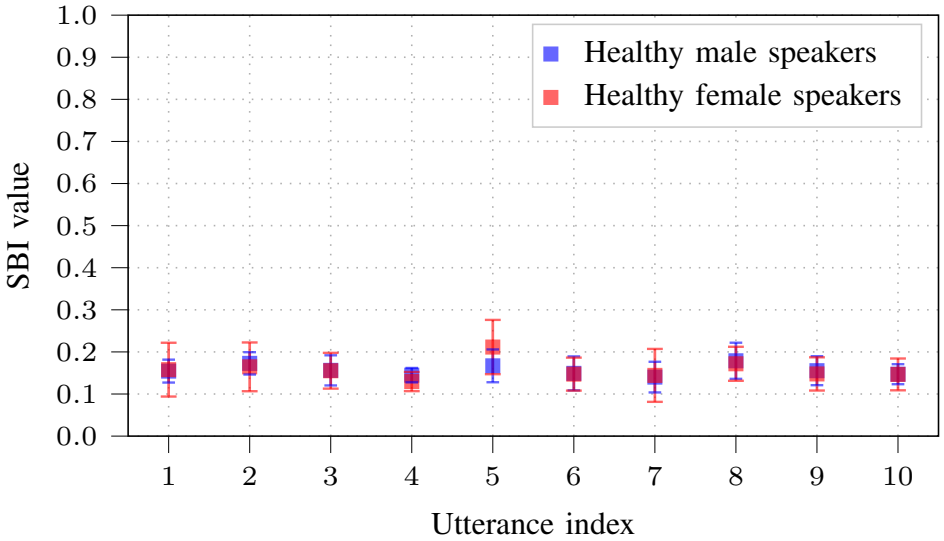
Robustness of SBI to gender and age variations

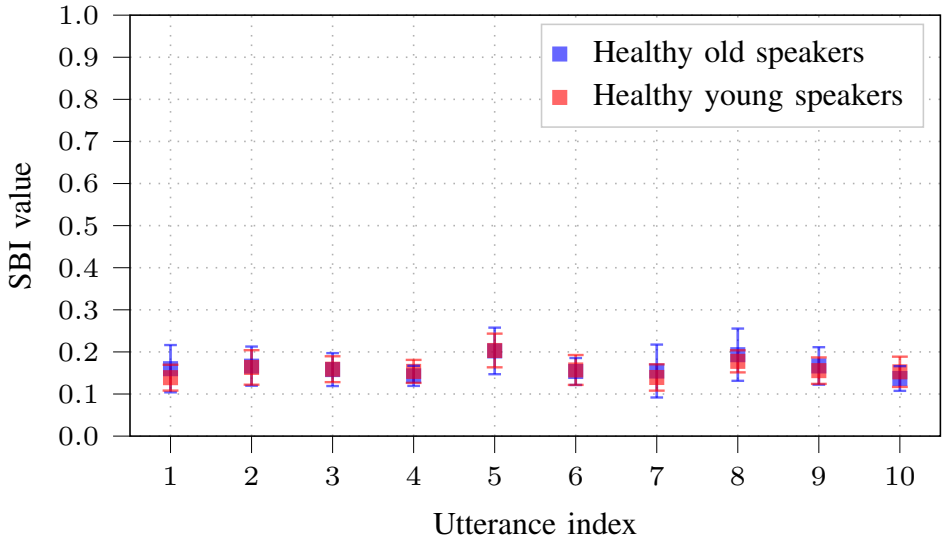
Investigating the robustness of the proposed SBI measure to the gender and age of speakers (only healthy)

- Comparing SBI value for **male** vs. **female** speakers
  - 20 (10 males and 10 females) speakers to represent the intelligible (reference) speech signals
  - 30 (15 males and 15 females) speakers to represent the test speech signals
- Comparing SBI value for **old** vs. **young** speakers
  - 18 (9 old and 9 young) speakers to represent the intelligible (reference) speech signals
  - 30 (15 old and 15 young) speakers to represent the test speech signals
- The random selection of disjoint subsets of reference and test speakers is repeated 100 times

Robustness of SBI to gender and age variations

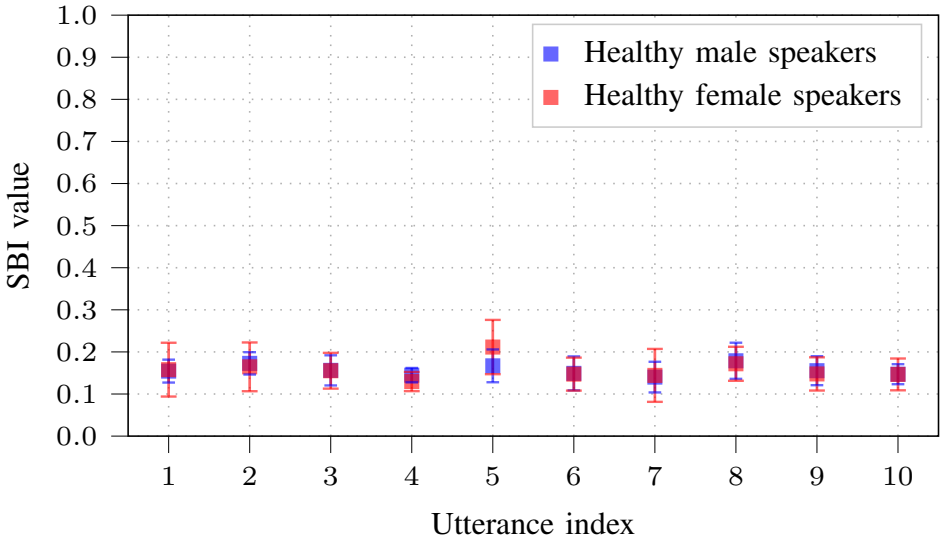
Investigating the robustness of the proposed SBI measure to the gender and age of speakers (only healthy)

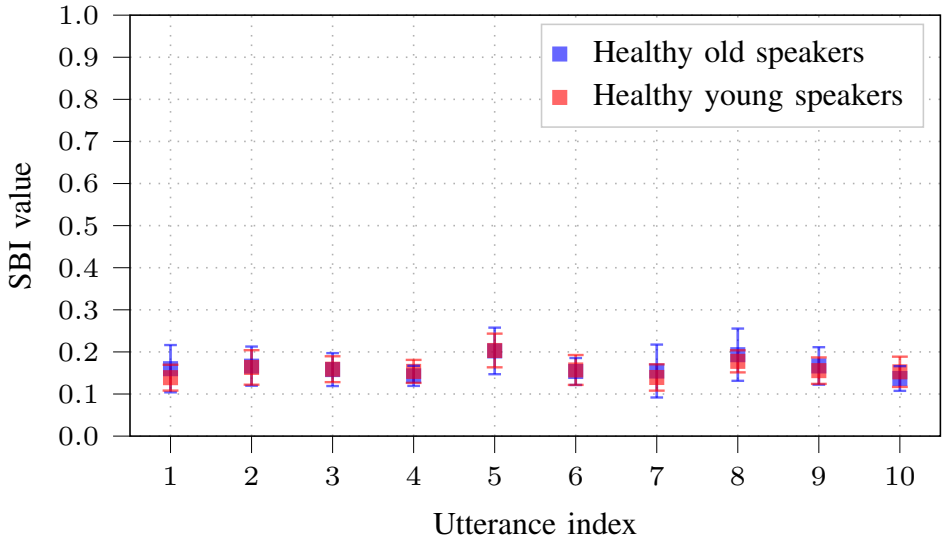




Robustness of SBI to gender and age variations

Investigating the robustness of the proposed SBI measure to the gender and age of speakers (only healthy)





- Mean SBI values are **similar** across the **two gender and age groups**, independently of the utterance
- Low mean SBI values ⇒ intelligible independently of the gender or age
- Independent t-test for each repetition ⇒ **not statistically significant difference** in the mean SBI values across old and young speakers
  ⇒ SBI is robust to age and gender

Outline

1. Automatic Pathological speech intelligibility assessment

2. Subspace-based pathological speech intelligibility assessment
   - Proposed SBI measure
   - Modulation spectrum and speech intelligibility
   - Empirical insights into SBI measure

3. Experimental results

4. Summary

Experimental results

- Dataset
  - English Universal Access database (UA-Speech) (Kim et al., 2008)
    - 763 isolated words from 15 CP patients and 13 healthy speakers
  - Dutch corpus of pathological and normal speech (COPAS) (Nuffelen et al., 2009)
    - 47 words from 16 HI patients and 22 healthy speakers
- State-of-the-art intelligibility measures
  - P-ESTOI (Janbakhshi et al., 2019a)
  - iVector- and ASR-based approach (Martínez et al., 2015)
- Evaluation (comparing subjective intelligibility vs. predicted scores)
  - Pearson correlation coefficient (R)
  - Spearman rank correlation coefficient (R_S)

Experimental results (considered scenarios)

i. Phonetically balanced scenarios
   - All speakers (healthy and pathological) utter exactly the same words

ii. Phonetically unbalanced scenarios
   - (i) Phonetic content within each group is the same but *completely different* across the two groups (227 words per speaker considering UA-Speech database)

Experimental results (considered scenarios)

i. Phonetically balanced scenarios
   - All speakers (healthy and pathological) utter exactly the same words

ii. Phonetically unbalanced scenarios
   - (i) Phonetic content within each group is the same but *completely different* across the two groups (227 words per speaker considering UA-Speech database)
   - (ii) Phonetic content within each group is the same but *partially different* across the two groups (304 words)
   - (iii) Phonetic content across all speakers is *partially different* (200 words)
   - (iv) Phonetic content across all speakers is *completely different* (16 words)

Random selection of subset of words is repeated 100 times

Experimental results (phonetically balanced scenarios)

$\{.\}^*$: non-significant correlations

	15 English CP patients		16 Dutch HI patients	
Measures	R	R_S	R	R_S
P-ESTOI	0.944	0.945	0.804	0.805
iVector	0.74	-	-	-
ASR	0.55	-	-	-
SBI	−0.856	−0.877	−0.480	-0.397^*
DSBI	−0.863	−0.934	−0.641	−0.603
MASBI	−0.821	−0.877	−0.682	−0.650

Experimental results (phonetically balanced scenarios)

$\{.\}^*$: non-significant correlations

	15 English CP patients		16 Dutch HI patients	
Measures	R	R_S	R	R_S
P-ESTOI	0.944	0.945	0.804	0.805
iVector	0.74	-	-	-
ASR	0.55	-	-	-
SBI	−0.856	−0.877	−0.480	−0.397*
DSBI	−0.863	−0.934	−0.641	−0.603
MASBI	−0.821	−0.877	−0.682	−0.650

- P-ESTOI ⇒ highest correlation values but is limited to only such phonetically balanced scenarios
- Proposed SBI, DSBI, and MASBI ⇒ high and significant correlations, outperforming iVector- and ASR-based approaches
- Incorporating short-time temporal information (i.e., as in the DSBI and MASBI) can yield improvement as opposed to SBI

Experimental results (phonetically unbalanced scenarios)

$\{.\}^*$: non-significant correlations

Measures	R	R_S
Phonetically unbalanced scenario		
SBI	-0.735 ± 0.028	-0.755 ± 0.038
DSBI	-0.699 ± 0.059	-0.731 ± 0.062
MASBI	-0.710 ± 0.060	-0.739 ± 0.073

Experimental results (phonetically unbalanced scenarios)

$\{.\}^*$: non-significant correlations

Measures	R	R_S
Phonetically unbalanced scenario		
SBI	-0.735 ± 0.028	-0.755 ± 0.038
DSBI	-0.699 ± 0.059	-0.731 ± 0.062
MASBI	-0.710 ± 0.060	-0.739 ± 0.073

- All measures $\Rightarrow$ typically high and significant correlations (applicable to these scenarios)
- Different phonetic content across speakers $\Rightarrow$ no improvements with incorporating short-time temporal information

Outline

1. Automatic Pathological speech intelligibility assessment

2. Subspace-based pathological speech intelligibility assessment
   - Proposed SBI measure
   - Modulation spectrum and speech intelligibility
   - Empirical insights into SBI measure

3. Experimental results

4. Summary

Summary

- Automatic pathological speech intelligibility SBI measure based on the assessment of the distance between subspaces spanned by dominant spectral patterns of intelligible (i.e., healthy) and pathological speech.
- SBI measure can capture pathology-induced distortions in the important spectral modulation cues.
- SBI is robust to gender- and age-induced changes.
- Two extensions of the SBI measure, i.e., the DSBI and MASBI measure
- The proposed measures outperform several non-blind state-of-the-art measures, and being also applicable to phonetically unbalanced scenarios.

Thank You

Reference

Elliott, T. M. and Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLOS Computational Biology*, 5(3):1–14.

Janbakhshi, P., Kodrasi, I., and Bourlard, H. (2019a). Pathological speech intelligibility assessment based on the short-time objective intelligibility measure. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Brighton , UK.

Janbakhshi, P., Kodrasi, I., and Bourlard, H. (2019b). Spectral subspace analysis for automatic assessment of pathological speech intelligibility. In *Proc. 20th Annual Conference of the International Speech Communication Association*, Graz, Austria.

Janbakhshi, P., Kodrasi, I., and Bourlard, H. (2020). Automatic pathological speech intelligibility assessment exploiting subspace-based analyses. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. In press.

Kim, H., Hasegawa-Johnson, M., Perlman, A., Gunderson, J., Huang, T., Watkin, K., and Frame, S. (2008). Dysarthric speech database for universal access research. In *Proc. 9th Annual Conference of the International Speech Communication Association*, pages 1741–1744, Brisbane, Australia.

Martínez, D., Lleida, E., Green, P., Christensen, H., Ortega, A., and Miguel, A. (2015). Intelligibility assessment and speech recognizer word accuracy rate prediction for dysarthric speakers in a factor analysis subspace. *ACM Transactions on Accessible Computing*, 6(3):10:1–10:21.

Nuffelen, G. V., Bodt, M. S. D., Middag, C., and Martens, J. P. (2009). Dutch corpus of pathological and normal speech (COPAS). Technical report, Antwerp University Hospital and Ghent University.

Orozco, J. R., Arias-Londoño, J. D., Vargas-Bonilla, J., González-Rátiva, M., and Noeth, E. (2014). New spanish speech corpus database for the analysis of people suffering from parkinson's disease. In *Proc. 9th International Conference on Language Resources and Evaluation*, pages 342–347, Reykjavik, Iceland.