# Effect of Acoustic Conditions on Algorithms to Detect Parkinson's Disease from Speech

**Juan Camilo Vásquez–Correa**, Joan Serrà–Julià, Juan Rafael Orozco–Arroyave, Elmar Nöth

Faculty of Engineering, University of Antioquia UdeA.
Pattern recognition Lab. Friedrich Alexander Universität. Erlangen-Nürnberg.
Telefónica Research.

*jcamilo.vasquez@udea.edu.co*




The 42nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP–2017)

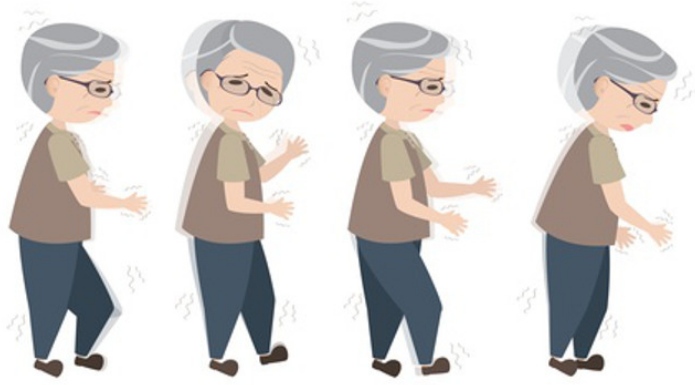
# Outline

UNIVERSIDAD DE ANTIOQUIA

# Introduction: Parkinson's Disease

UNIVERSIDAD DE ANTIOQUIA

- Second most prevalent neurological disorder worldwide.
- Patients develop several motor and non-motor impairments.
- Speech impairments are one of the earliest manifestations.

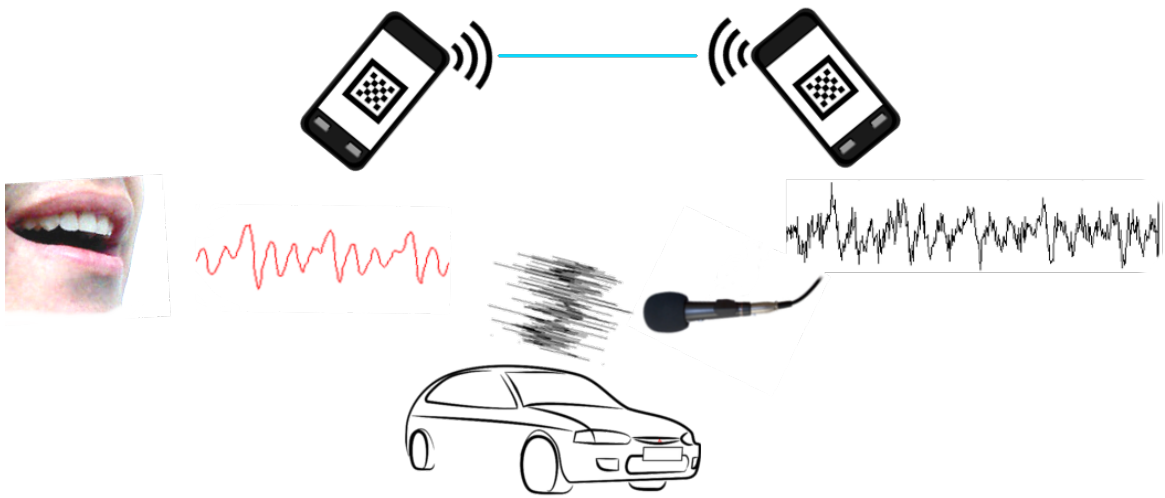


# Introduction: Motivation

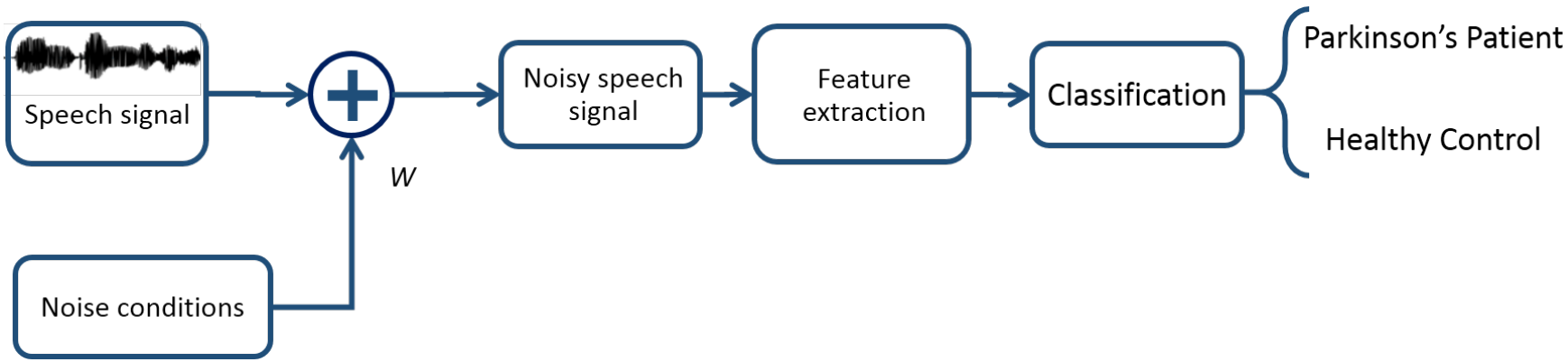
UNIVERSIDAD DE ANTIOQUIA

- Several state of art algorithms.
- Comparison under the same data and conditions.
- Performance under real world noisy conditions is unknown.
  - Background noise (Street, car, cafeteria...)
  - Distortion
  - Telephone channels

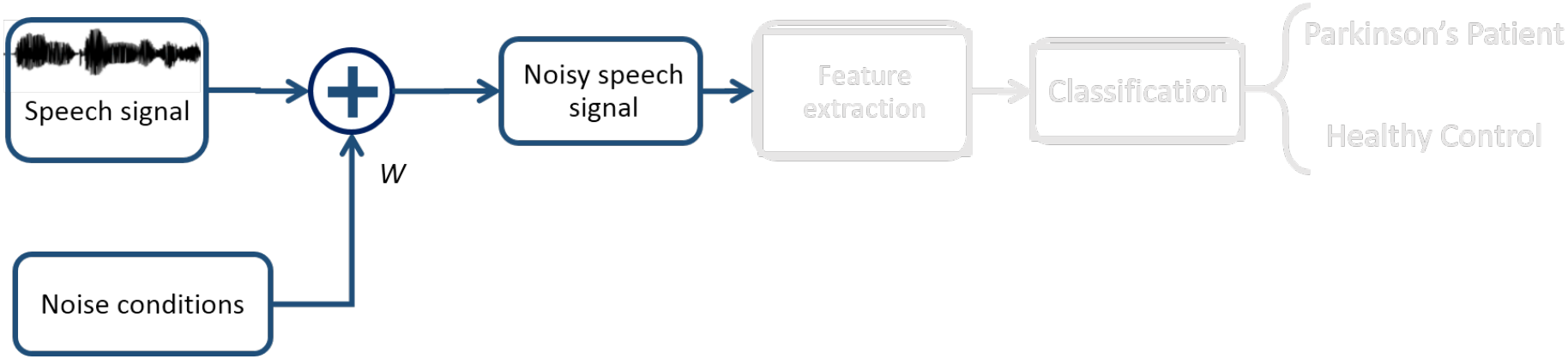
**What happen if the audio quality is degraded?**



# Methodology



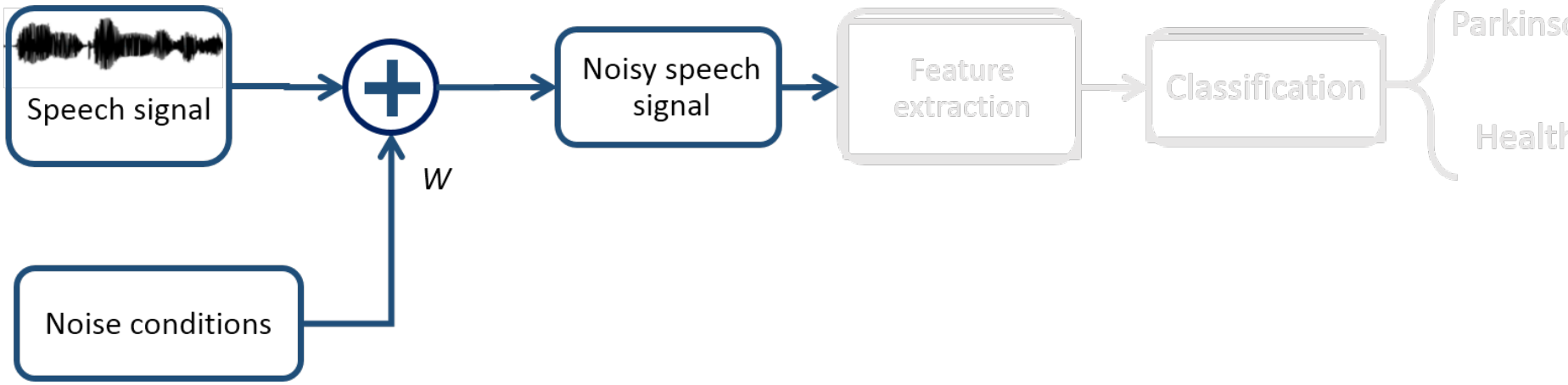
# Methodology: Non-Controlled Acoustic Conditions



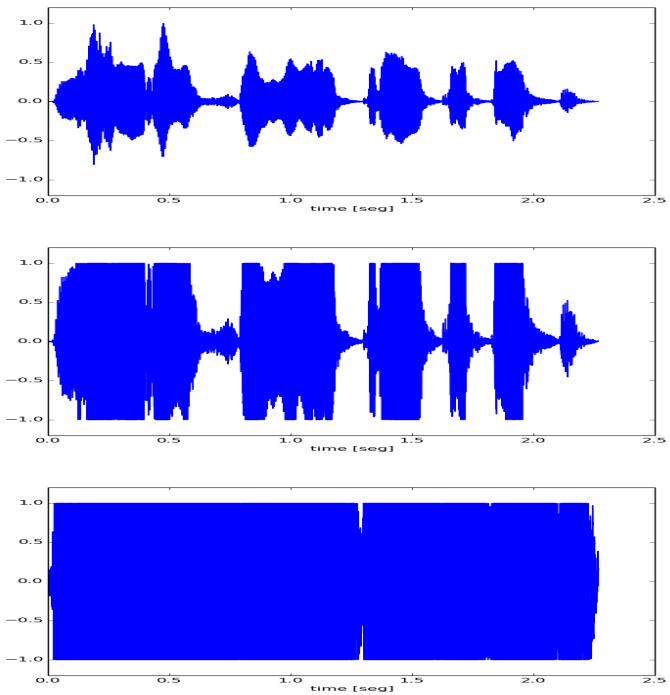

- Distortion
- Dynamic Compression
- Background noise
- Audio codecs
- Telephone Channels

# Methodology: Non-Controlled Acoustic Conditions



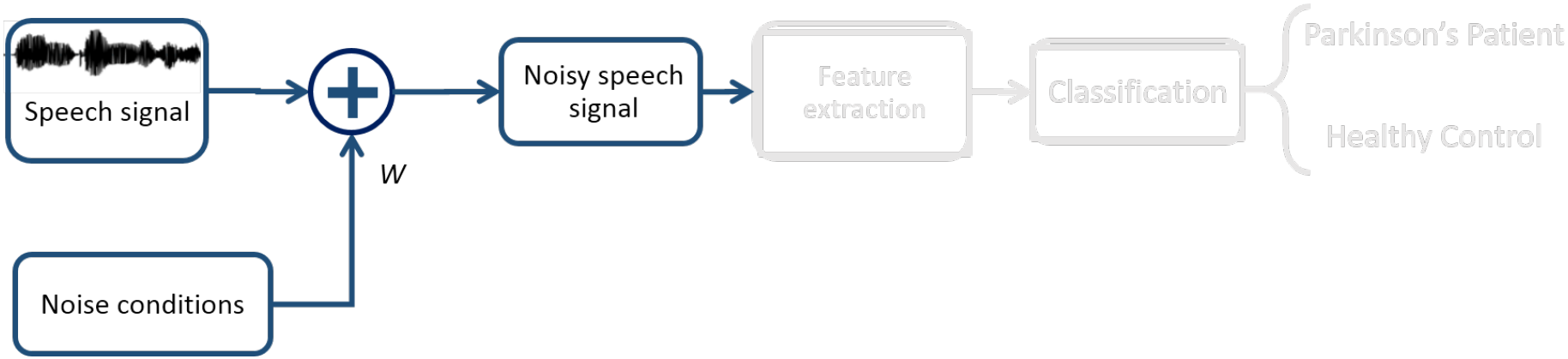



- Distortion
- Dynamic Compression
- Background noise
- Audio codecs
- Telephone Channels

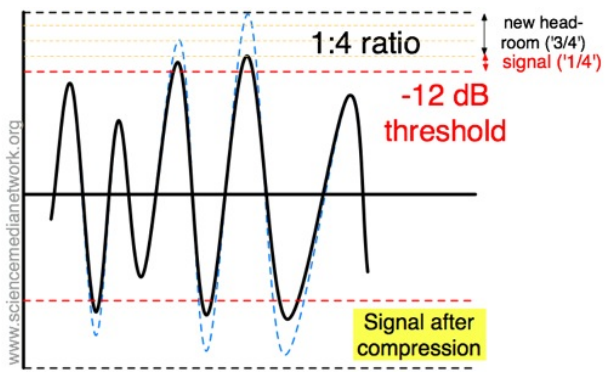


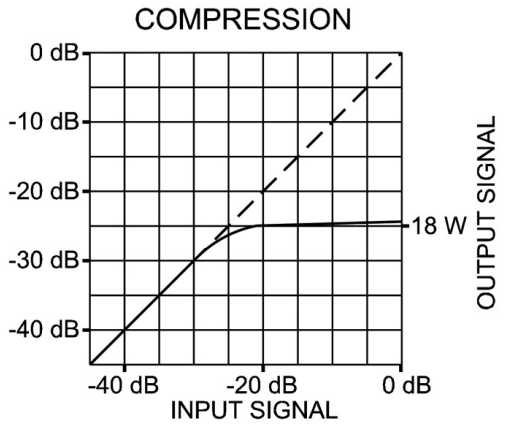
# Methodology: Non-Controlled Acoustic Conditions





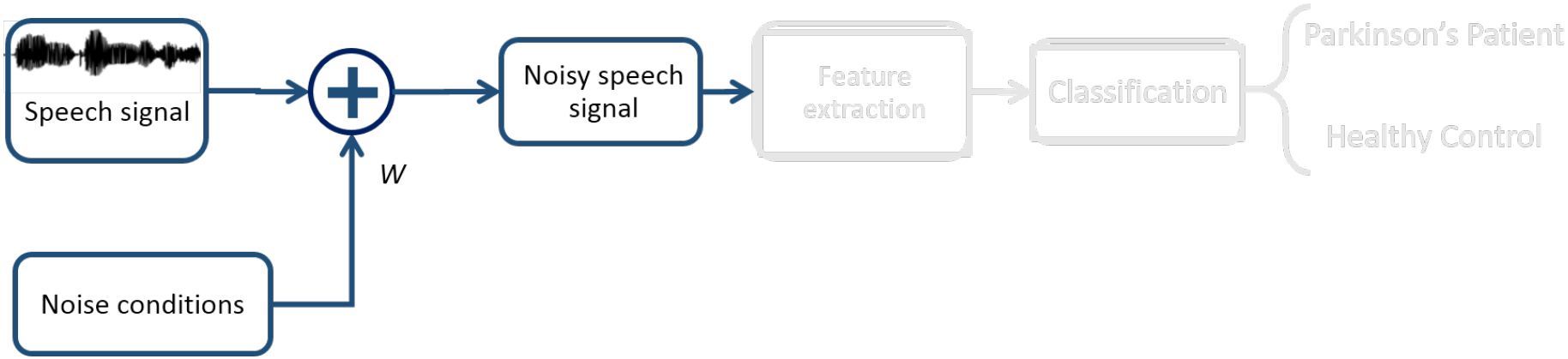

- Distortion
- Dynamic Compression
- Background noise
- Audio codecs
- Telephone Channels



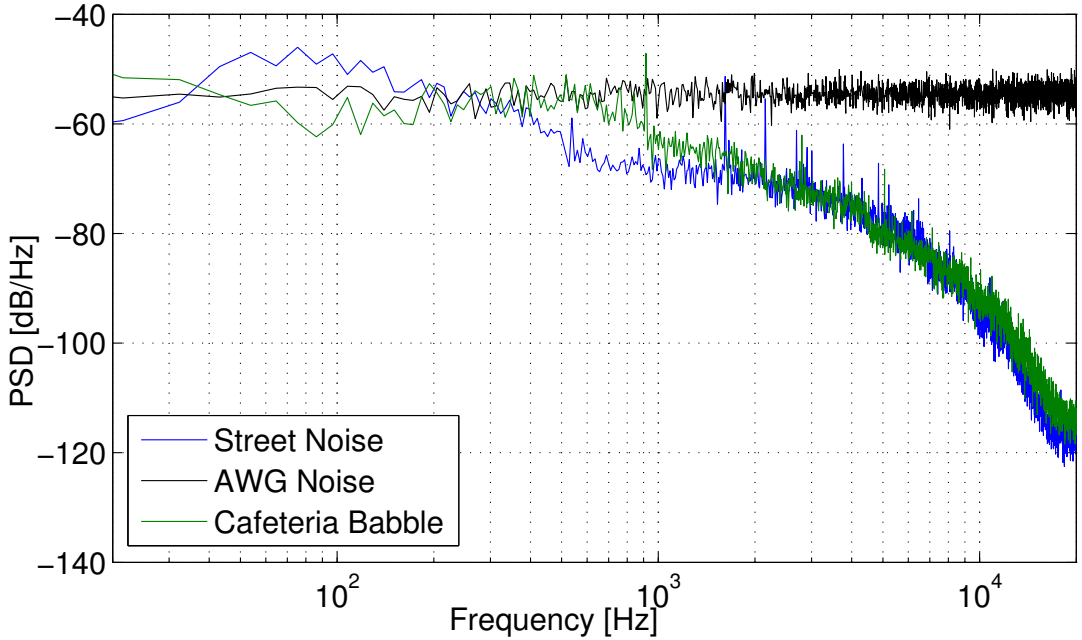



# Methodology: Non-Controlled Acoustic Conditions



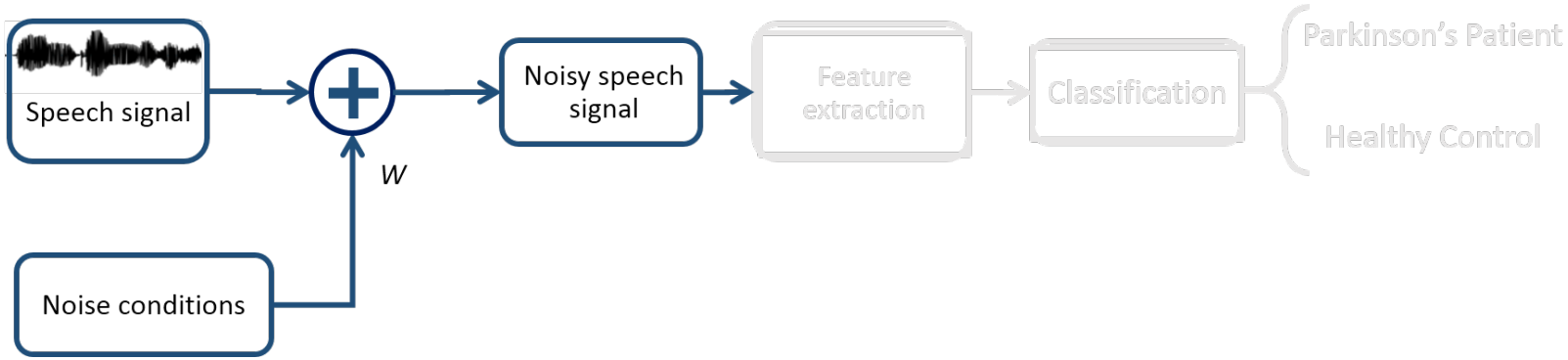



- Distortion
- Dynamic Compression
- Background noise
- Audio codecs
- Telephone Channels



# Methodology: Non-Controlled Acoustic Conditions

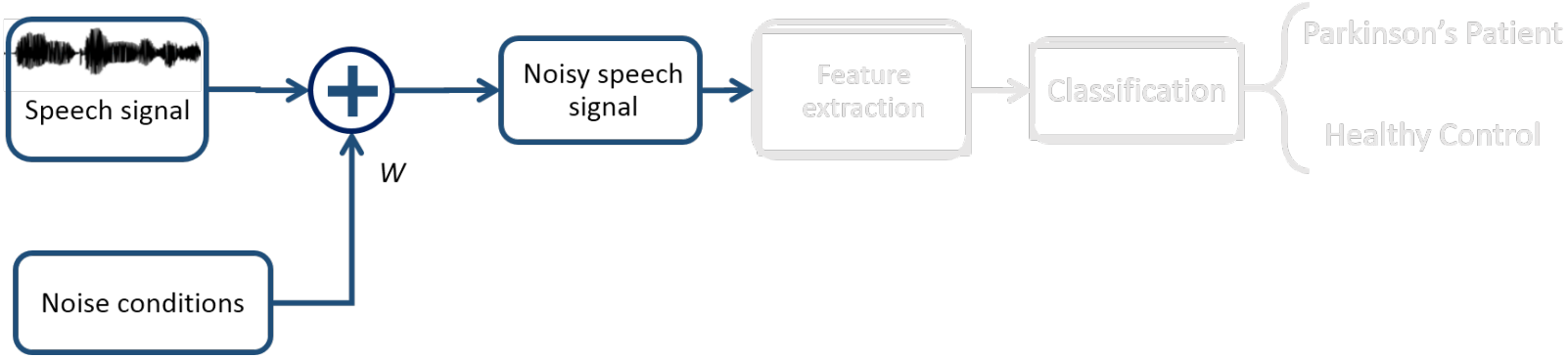





- Distortion
- Dynamic Compression
- Background noise
- Audio codecs
- Telephone Channels

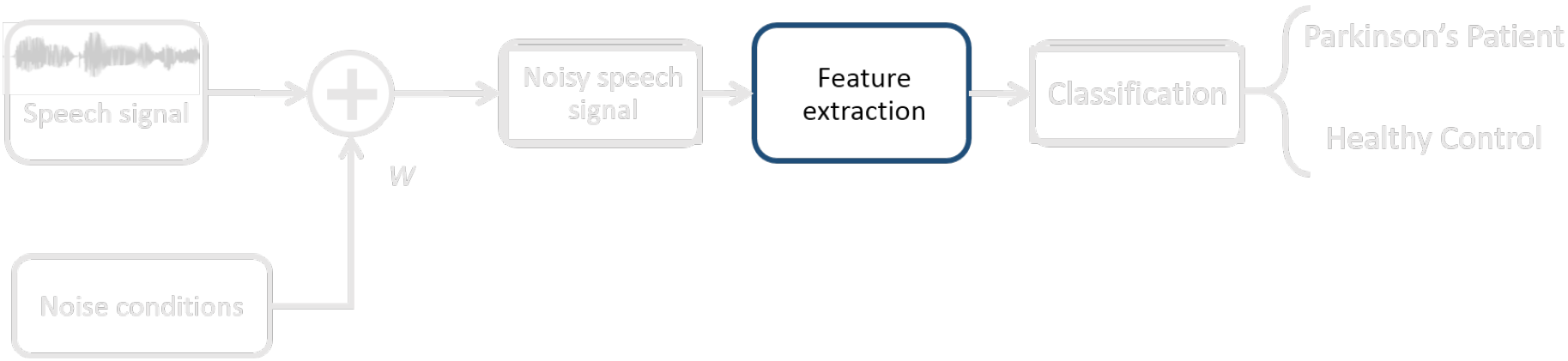
# Methodology: Non-Controlled Acoustic Conditions





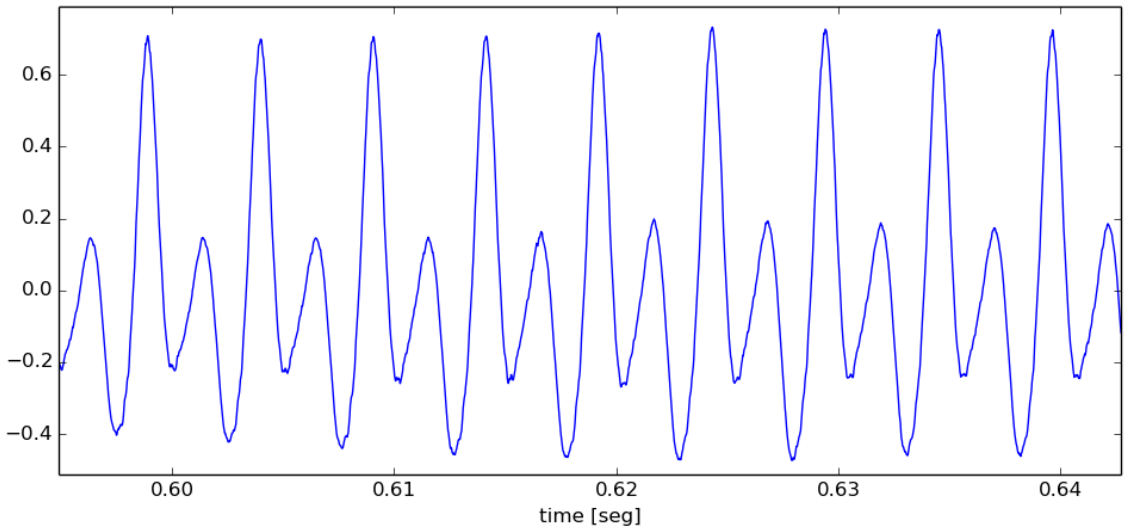

- Distortion
- Dynamic Compression
- Background noise
- Audio codecs
- Telephone Channels

# Methodology: Feature Extraction




- Phonation analysis in sustained vowels
- Voiced/Unvoiced modeling
- Articulation in Voiced/Unvoiced transitions
- Gaussian mixture models - supervectors
- OpenSMILE :)

# Feature Extraction: Phonation Analysis[1]



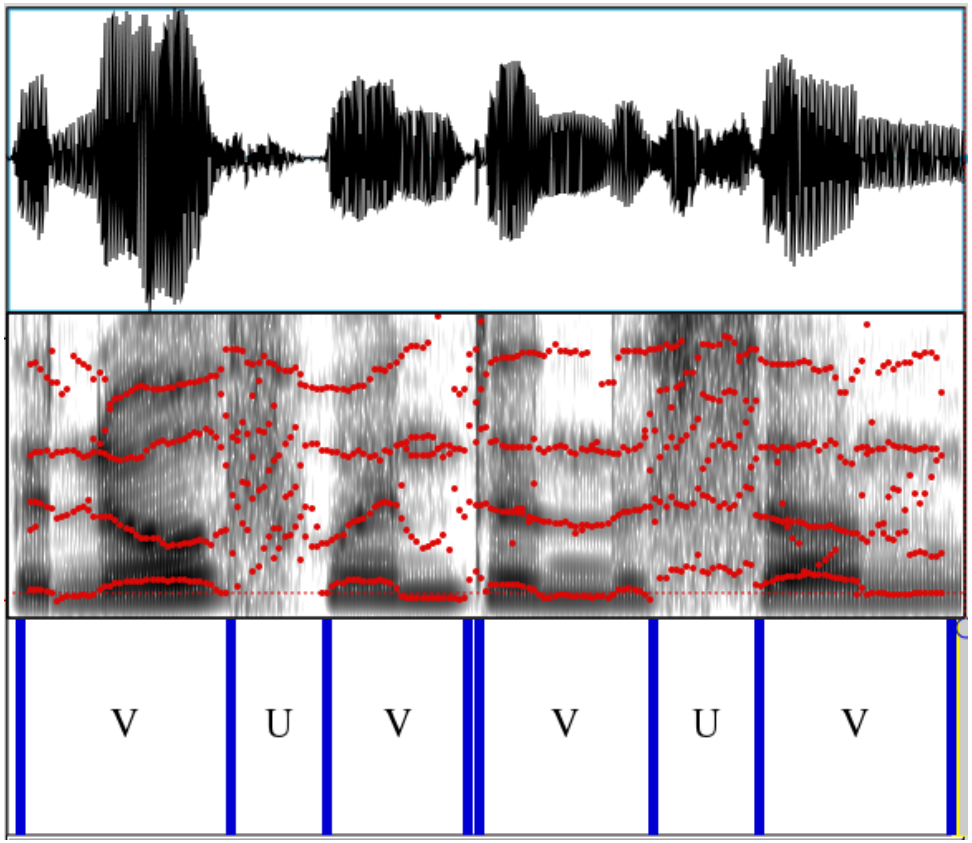

- Fundamental frequency $F_0$
- Jitter
- Shimmer
- Energy
- Amplitude perturb. quotient
- Pitch perturb. quotient

[1] J. R. Orozco-Arroyave, E. A. Belalcazar-Bolaños, et al. "Characterization Methods for the Detection of Multiple Voice Disorders: Neurological, Functional, and Laryngeal Diseases". In: *IEEE Journal of Biomedical and Health Informatics* 19.6 (2015), pp. 1820–1828.

# Feature Extraction: Voiced/Unvoiced Modeling[2]

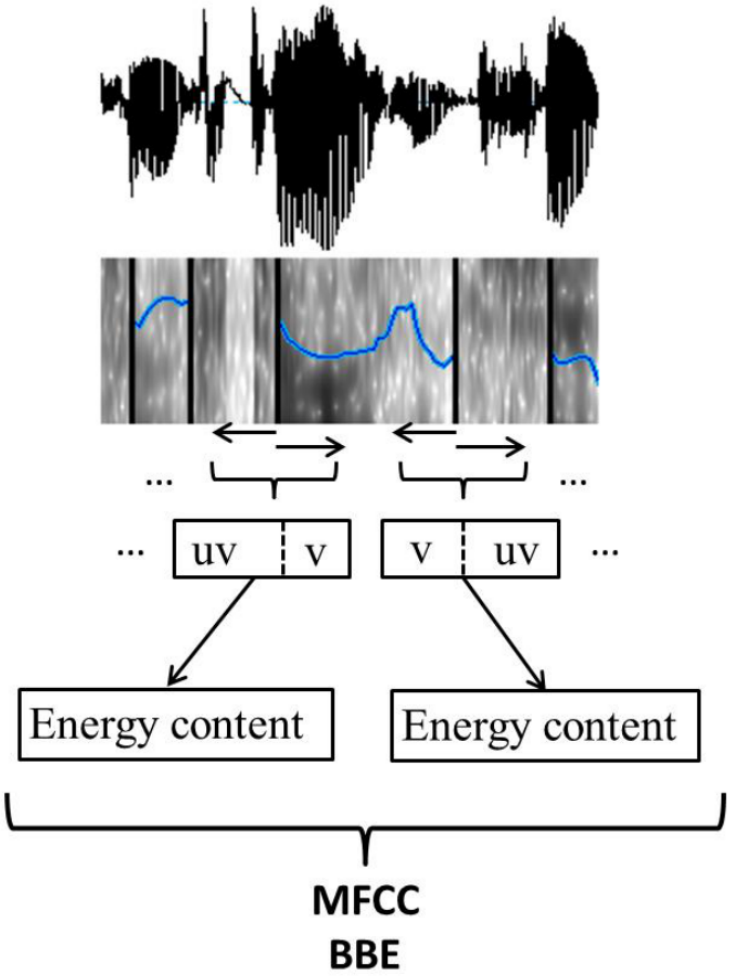
UNIVERSIDAD DE ANTIOQUIA

- Voiced: MFCC, jitter, shimmer, formant freq., $F_0$, energy.
- Unvoiced: MFCC, energy in Bark scale




[2] J. R. Orozco-Arroyave, F. Hönig, S. Skodda, et al. "Automatic detection of Parkinson's disease from words uttered in three different languages.". In: *15th Anual Conference of the Speech and Communication Association (INTERSPEECH)*. 2014, pp. 1573–1577.

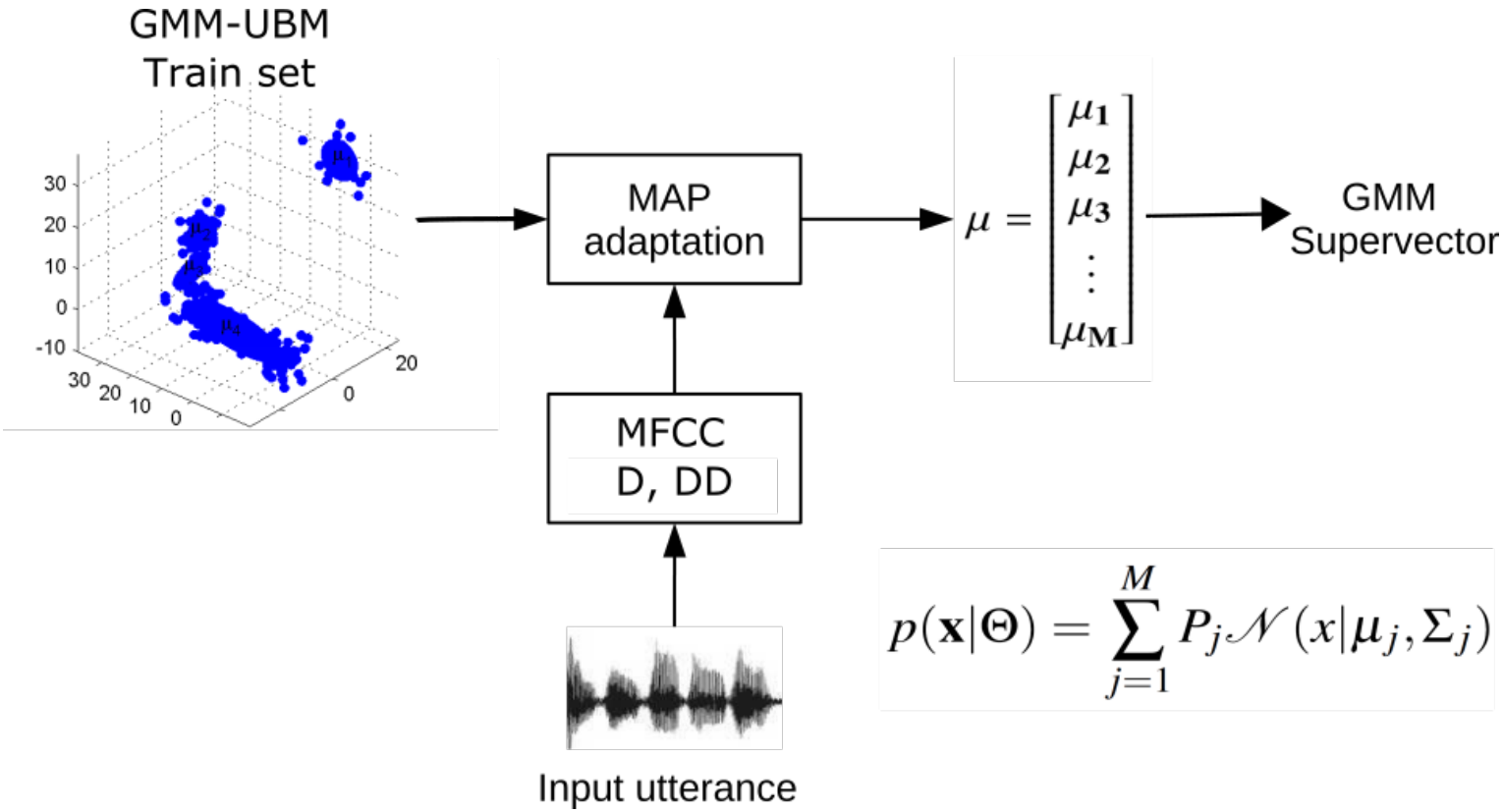
# Feature Extraction: Voiced/Unvoiced Transition[3]




The hypothesis: the difficulty of the patients to start/stop walking is also reflected in the process to start/stop the vocal fold vibration

[3] J. R. Orozco-Arroyave, F. Hönig, J. F. Vargas-Bonilla, et al. "Voiced/unvoiced transitions in speech as a potential bio–marker to detect Parkinson's disease". In: *16th Anual Conference of the Speech and Communication Association (INTERSPEECH)*. 2015, pp. 95–99.

# Feature Extraction:GMM Supervector[4]




[4] Tobias Bocklet et al. "Automatic evaluation of Parkinson's speech-acoustic, prosodic and voice related cues". In: *14th Anual Conference of the Speech and Communication Association (INTERSPEECH)*. 2013, pp. 1149–1153.

# Feature Extraction: OpenSMILE :)[5]

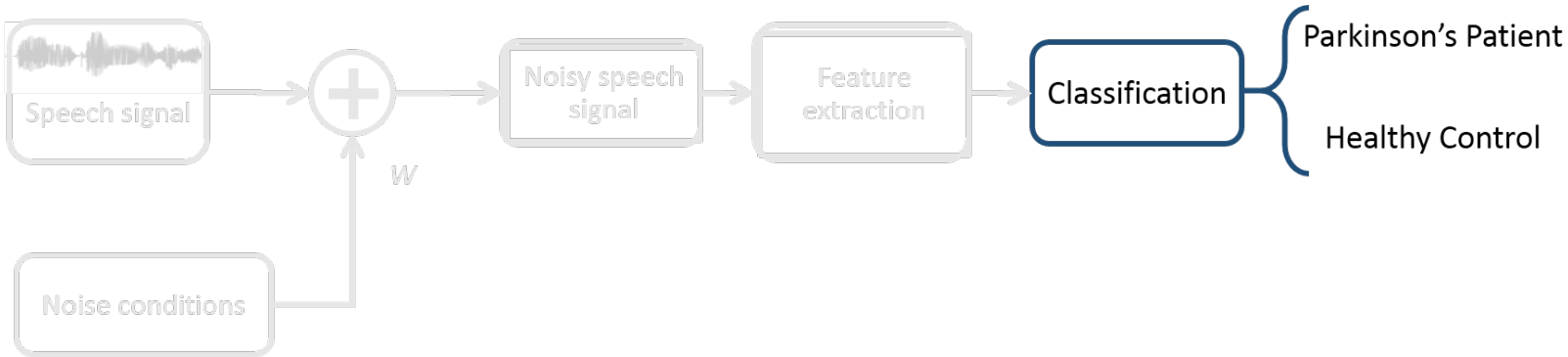
UNIVERSIDAD DE ANTIOQUIA

openSMILE:)
by audEERING™

- Standard toolkit for speech processing
- 6373 static acoustic features
  - Energy
  - $F_0$
  - MFCC
  - Duration ...

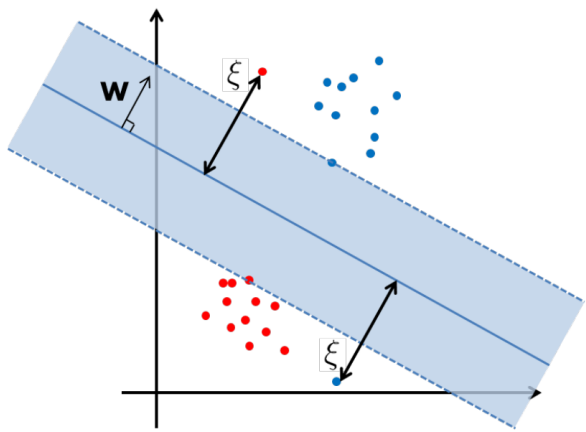
[5] Florian Eyben and Björn Schuller. "openSMILE:): the Munich open-source large-scale multimedia feature extractor". In: *ACM SIGMultimedia Records* 6.4 (2015), pp. 4–13.

# Methodology: Classification




- Support vector machine
- RBF/Linear kernel
- Leave one out CV
- Accuracy as performance measure

$$\underset{\mathbf{w},b,\xi_i}{\text{minimize}} \quad \frac{1}{2}||\mathbf{w}||^2 + C\left(\sum_{i=1}^{M}\xi_i\right)^k$$



# Methodology: Data[6]

UNIVERSIDAD DE ANTIOQUIA

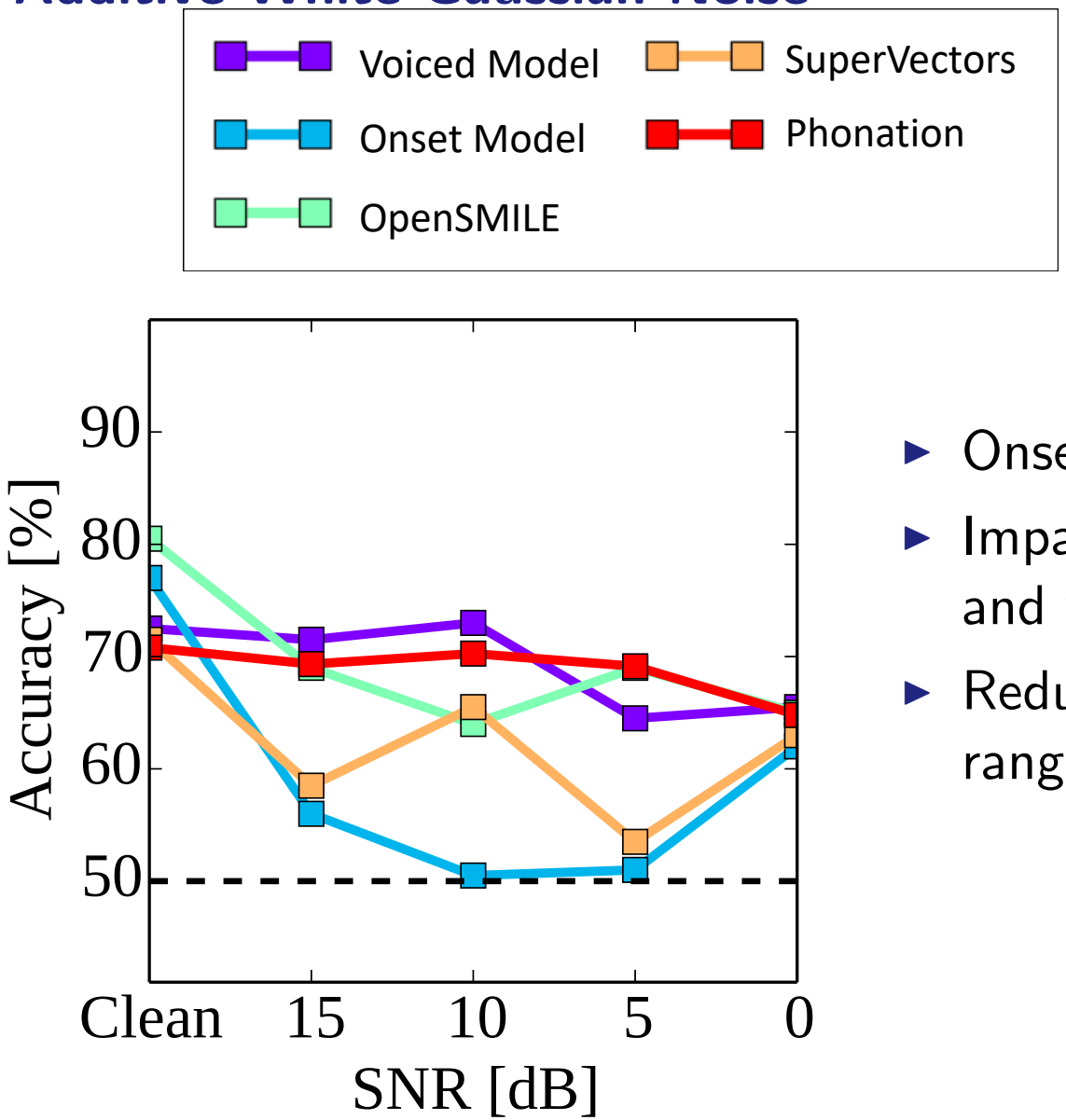
- 100 native Spanish speakers (Colombians): 50 Parkinson's patients & 50 Healthy controls.
- Recorded in a sound-proof booth with professional equipment.
- Different speech tasks.
  - Sustained vowels
  - Readtext
  - Monologue

[6] J. R. Orozco-Arroyave, Vargas-Bonilla J. F., et al. "New spanish speech corpus database for the analysis of people suffering from Parkinson's disease.". In: *9th Language Resources and Evaluation Conference, (LREC)*. 2014, pp. 342–347.

# Results: Matched Conditions

UNIVERSIDAD DE ANTIOQUIA

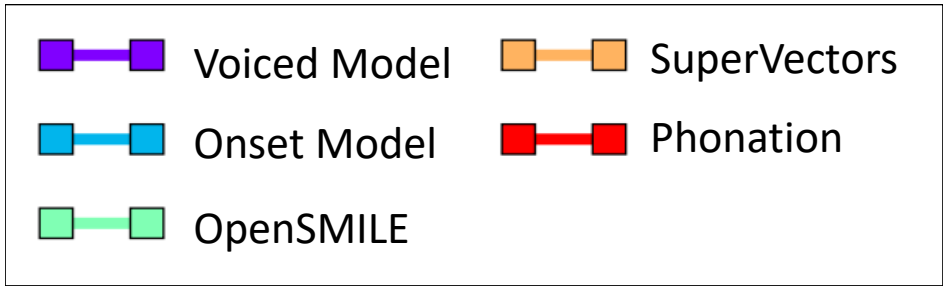
## Additive White Gaussian Noise

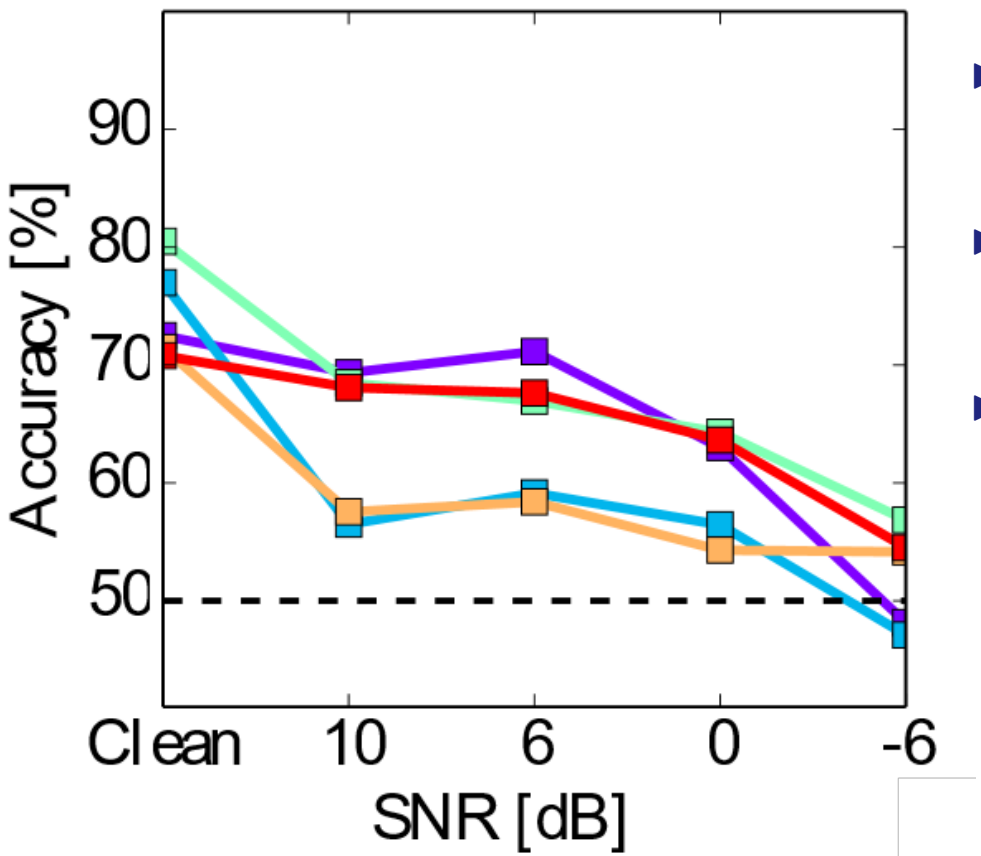



- Onset is the most affected
- Impact reduced in vowels and voiced
- Reduction in performance ranges from 10% to 20%

UNIVERSIDAD DE ANTIOQUIA

## Environmental Noise




- After SNR=6 dB high reduction
- The effect is similar for all algorithms
- Onset, offset and SuperVectors are the most affected

UNIVERSIDAD DE ANTIOQUIA

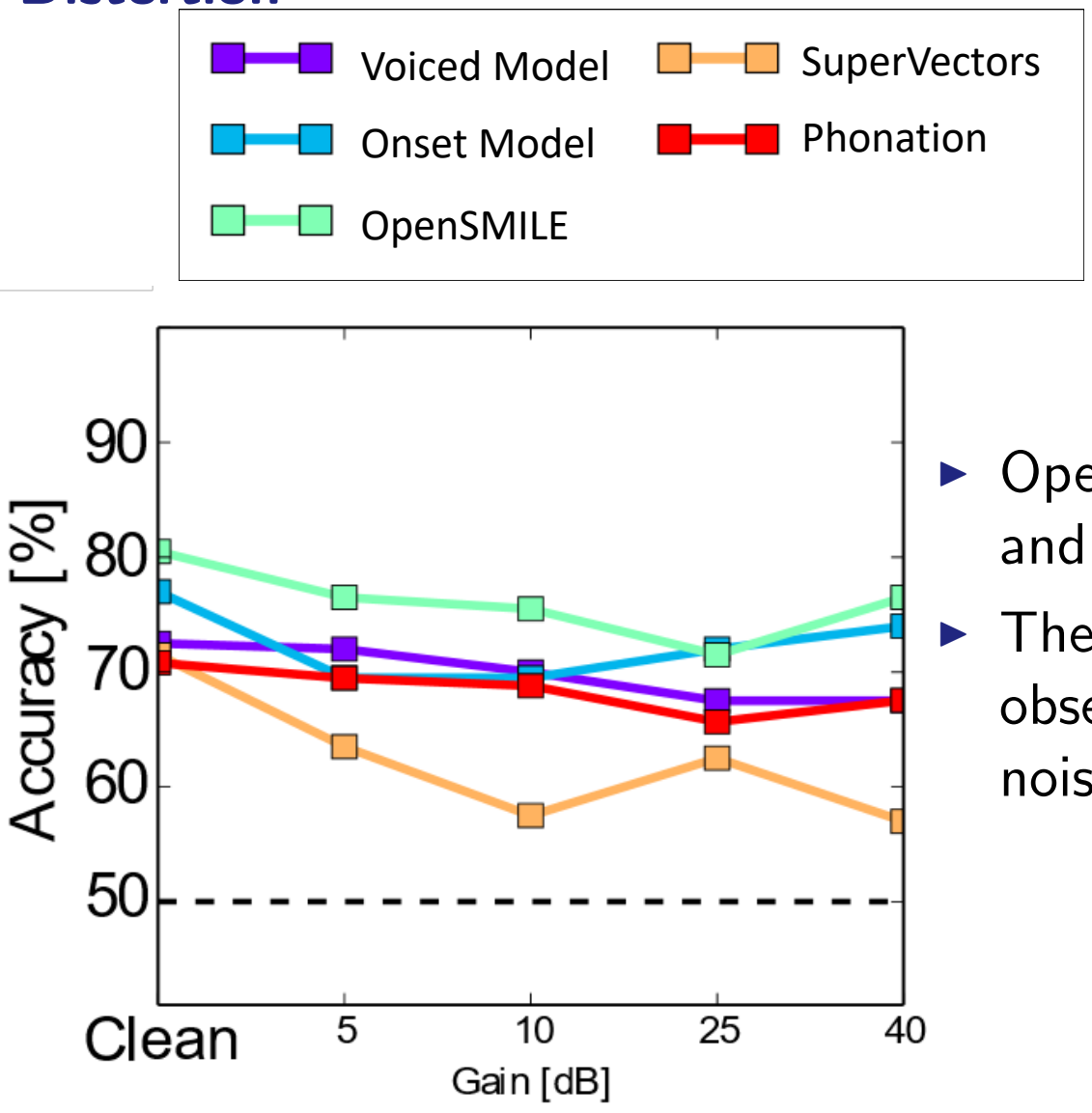
## Environmental Noise (2)

- Reverberated room noise causes the most performance reduction.
- Street and car noises have lower impact over the clean conditions.
- There is not a significance difference among the different noises relative to AWGN.

# Results: Matched Conditions

UNIVERSIDAD DE ANTIOQUIA

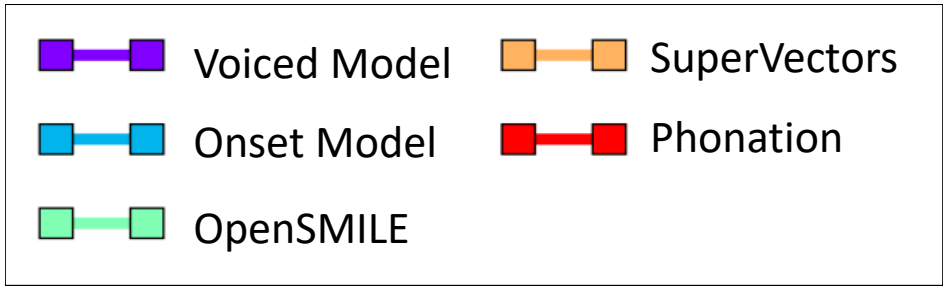
## Distortion

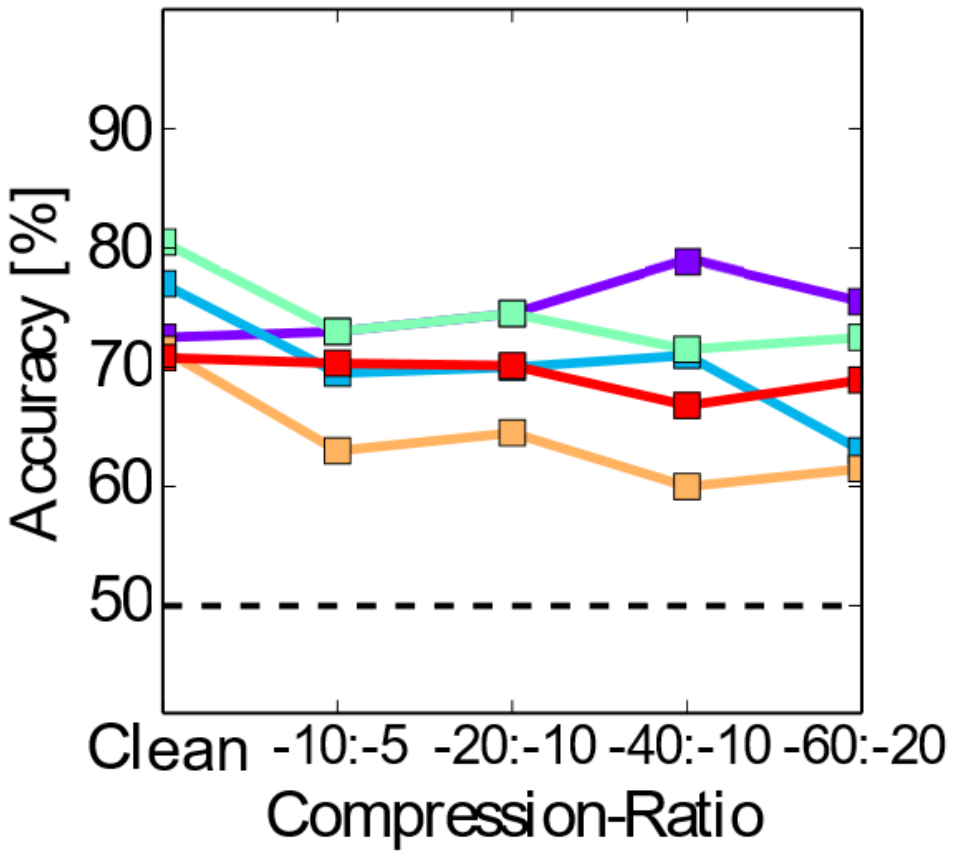



- OpenSMILE, vowels, voiced and onset are not affected
- The effect is lower than the observed for background noise

UNIVERSIDAD DE ANTIOQUIA

## Compression




- Some compression ratios may improve the results (voiced)

## Codecs

| | Vowels | Voiced | Onset | Offset | OS | SV |
|---|---|---|---|---|---|---|
| Clean | 72% | 74% | 82% | 81% | 81% | 72% |
| Opus | 74% | 79% | 86% | 80% | 87% | 69% |
| Silk | 71% | 75% | 75% | 75% | 75% | 61% |
| A-law | 75% | 78% | 73% | 67% | 64% | 62% |
| G.722 | 74% | 82% | 87% | 75% | 79% | 63% |
| GSM-FR | 73% | 82% | 70% | 64% | 76% | 68% |

- Opus and G.722 generally improve the results
- GSM improves the results for vowels and voiced

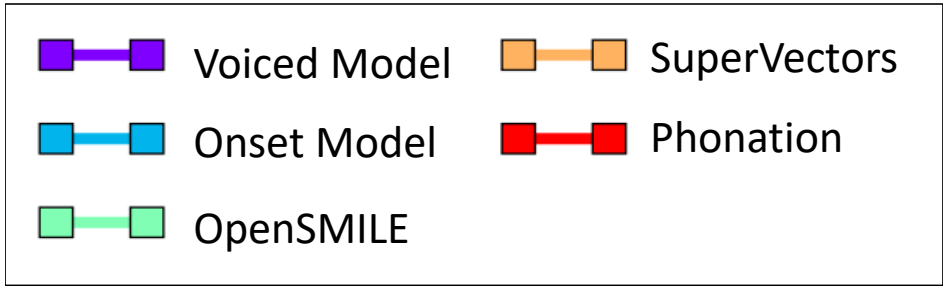
UNIVERSIDAD DE ANTIOQUIA

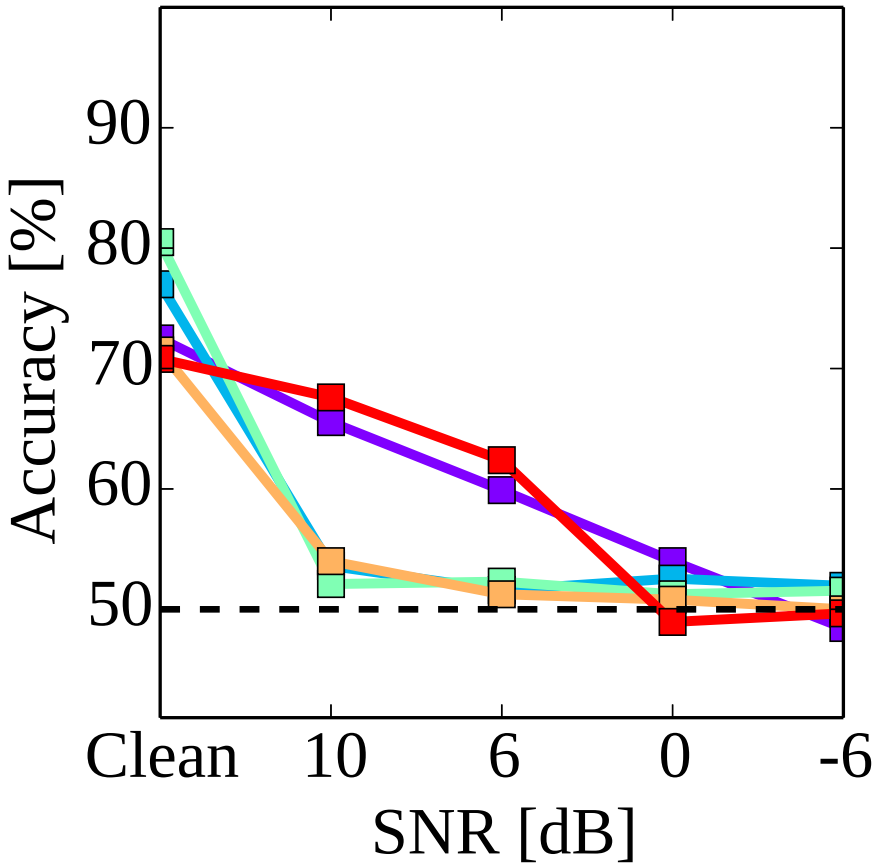
## Channels

| | Vowels | Voiced | Onset | Offset | OS | SV |
|---|---|---|---|---|---|---|
| Clean | 72% | 74% | 82% | 81% | 81% | 72% |
| Hangouts | 76% | 76% | 79% | 67% | 85% | 64% |
| Skype | 76% | 73% | 61% | 71% | 79% | 71% |
| Landline | 75% | 75% | 66% | 67% | 78% | 75% |
| Mobile | 73% | 76% | 65% | 57% | 76% | 71% |

- Hangouts and landline generally improve the results
- Mobile is the most affected, specially for onset & offset

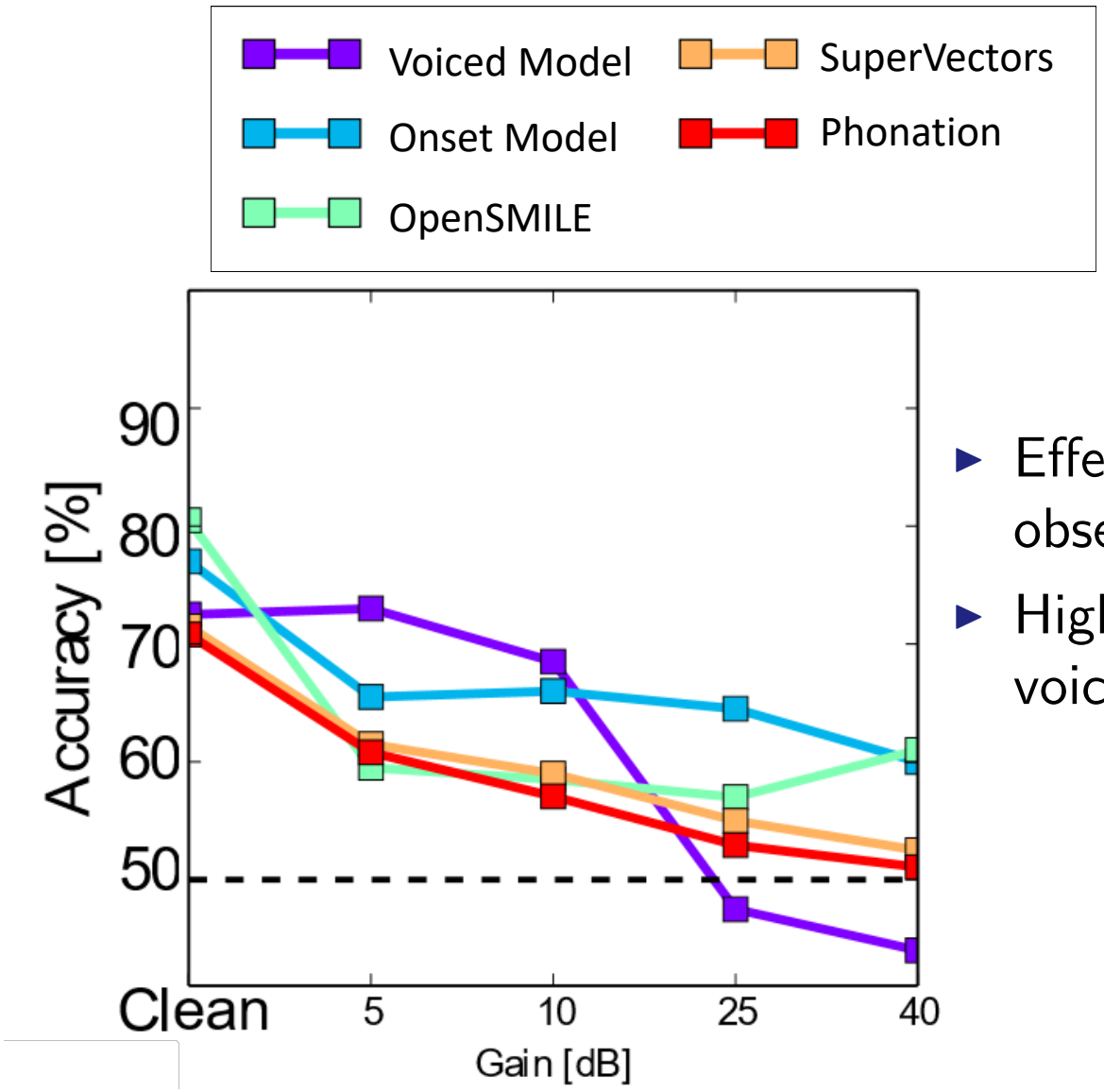
# Results: Mismatched Background Noise




- The effect is more **critical** than matched.
- Voiced and Vowels are the less affected
- Onset, OpenSMILE and SuperVectors are the most affected

# Results: Mismatched Distortion




- Effect of distortion is observed
- High impact for vowels and voiced.

# Conclusion

UNIVERSIDAD DE ANTIOQUIA

- Results in clean conditions range from 70% to 85%.

# Conclusion



- Results in clean conditions range from 70% to 85%.
- Results do not always decrease monotonically relative to the noise level.

# Conclusion



- Results in clean conditions range from 70% to 85%.
- Results do not always decrease monotonically relative to the noise level.
- Background noise is the most critical condition.

# Conclusion



- Results in clean conditions range from 70% to 85%.
- Results do not always decrease monotonically relative to the noise level.
- Background noise is the most critical condition.
- Audio codecs and dynamic compression can improve the results.

# Conclusion



- Results in clean conditions range from 70% to 85%.
- Results do not always decrease monotonically relative to the noise level.
- Background noise is the most critical condition.
- Audio codecs and dynamic compression can improve the results.

- The effect produced by telephone channels is not too critical.

# Conclusion



- Results in clean conditions range from 70% to 85%.
- Results do not always decrease monotonically relative to the noise level.
- Background noise is the most critical condition.
- Audio codecs and dynamic compression can improve the results.

- The effect produced by telephone channels is not too critical.
- Mismatched conditions is a problem which need to be solved. (Data augmentation, speech enhancement).

# Effect of Acoustic Conditions on Algorithms to Detect Parkinson's Disease from Speech

**Juan Camilo Vásquez–Correa**, Joan Serrà–Julià, Juan Rafael Orozco–Arroyave, Elmar Nöth

Faculty of Engineering, University of Antioquia UdeA.
Pattern recognition Lab. Friedrich Alexander Universität. Erlangen-Nürnberg.
Telefónica Research.

*jcamilo.vasquez@udea.edu.co*




The 42nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP–2017)