



Acute stress detection from voice as an add-on security measure in ATM and airport operations

Milan Rusko, Marián Trnka, Sakhia Darjaa, Marian Ritomský

Institute of Informatics of the Slovak Academy of Sciences, Bratislava (UI SAV)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 832969.



Voice communication in ATC and airports

- Voice communication systems (VCS) enable its users to initiate, receive, attend to and maintain communication over radio or telephone.
- We focus on the ones used by the **air control centers**, both en route and on approach, and by **airport control towers** to communicate with pilots
- They can be divided in:
 - Ground/Air (G/A) voice communications, or radio communications, **between air traffic controllers and aircraft pilots**, and
 - Ground/Ground (G/G) voice communications, radio or telephone communications **between air traffic controllers and pilots** for the coordination of operations, and **between air traffic controllers and support staff** for management purposes.

Stress in air traffic, ATM and airports

- **Passenger stress**
- **Common operational stress of employees under standard conditions**
- **Acute stress in emergency situations**

We focus mainly on ATCo-s and pilots

We focus on identifying potential emergency situations mainly from ATCo-s and pilots voices)

We try to detect stress and the relating emotions , such as hot-anger, fear and anxiety, but also cold-anger, sadness, depression, etc.

Stress

- **Defining of stress** is a notoriously difficult problem. According to Oxford English dictionary, one of the meanings of the term “stress” is: A state of mental or emotional strain or tension resulting from adverse or demanding circumstances. (these can be emergency situations)
- Stressors in ATM: time stress, high cognitive workload, multitasking conditions, sleep deprivation, frustration over contradictory information, in emergency situations psychological tension, and even pain, and other.

Stressor order	Description	Stressors
0	Physical	Vibration Acceleration (G-force) Personal equipment, Pressure Breathing, Breathing gas mixture
1	Physiological	Medicines, Narcotics, Alcohol nicotine, Fatigue, Sleep deprivation Dehydration, Illness, Local anaesthetic
2	Perceptual	Noise (Lombard effect), Poor communication channel, Listener has poor grasp of the language
3	Psychological	Emotion, Workload Task-related anxiety Background anxiety

[Hansen 2000]

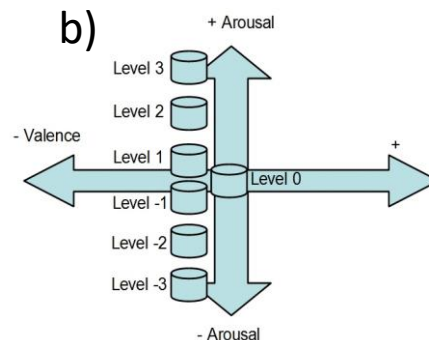
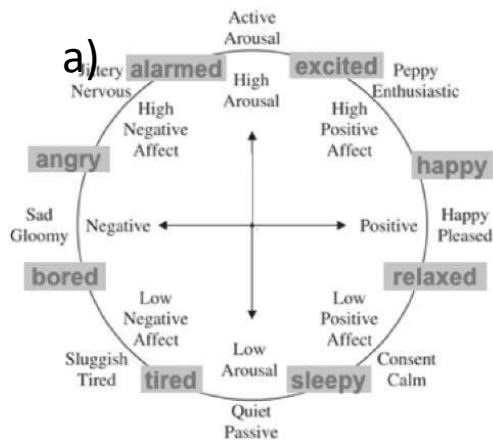
“Stress level evaluation from speech” task in SATIE project – main issues

- **Signal quality is low** in ATS-Pilot radio communication
- **Utterances are too short**
- **Various mother-tongues and various cultures**
- **Speech characteristics** reflecting stress **are highly speaker dependent** (personality, culture, mental and physical state etc.)
- **Stress cues in voice are highly non- specific**
(similar to those reflecting state of health, emotions, moods, or even personality features.)
- **Nearly no databases are available**, real data hard to obtain due to ethical reasons.
- **Most available databases are acted.** No real-world data available.

Relation between stress and emotions:

Dimensional versus categorical models of affective states

- a) Emotion wheel shows the position of emotion categories in 2 dimensional model
- b) The Crisis database structure – Emotional space sampling



Shott argued that to experience emotion, **people first experience physiological arousal and then they label this arousal as emotion**. [Shott S. Emotion and social life: a symbolic interactionist analysis. *Am J Sociol* 1979;84:1317–34.]

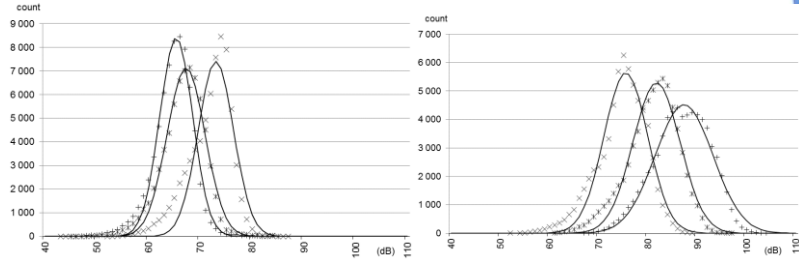
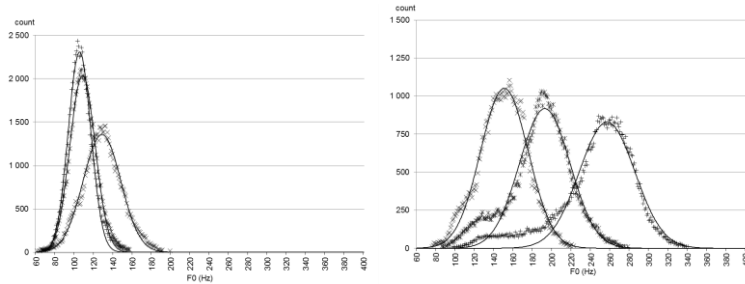
Stress cues in speech – F0 and Intensity

F0 histograms of speaker MR for decreasing (a) and increasing levels of arousal (b) Intensity histograms of speaker MR for decreasing (c) and increasing levels of arousal (b) a)

(b)

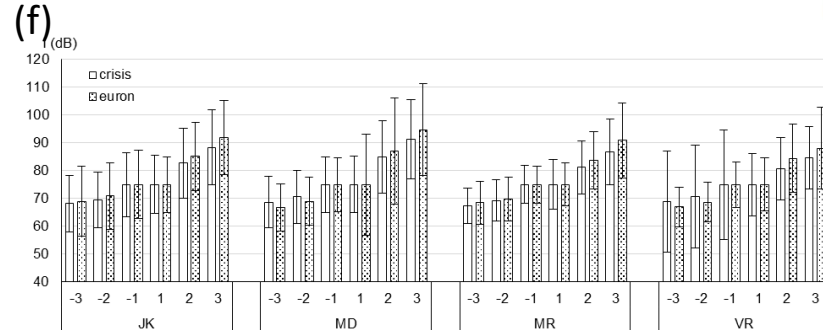
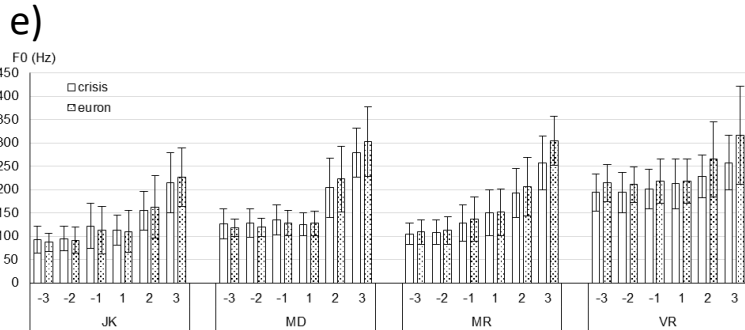
(c)

(d)



F0 mean and variance for four speakers, for six levels of arousal in increasing order (e)

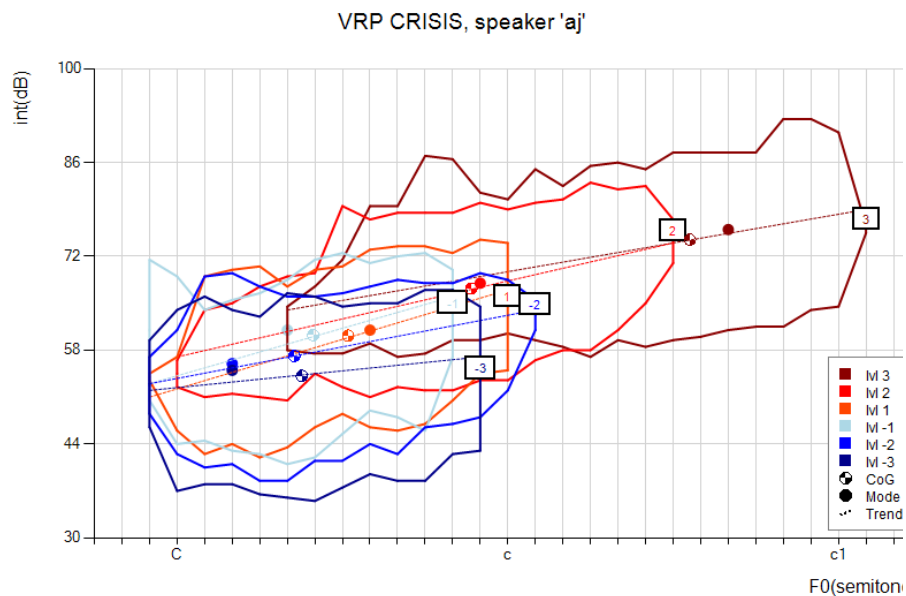
Intensity mean and variance for four speakers for six levels of arousal in increasing order (f)



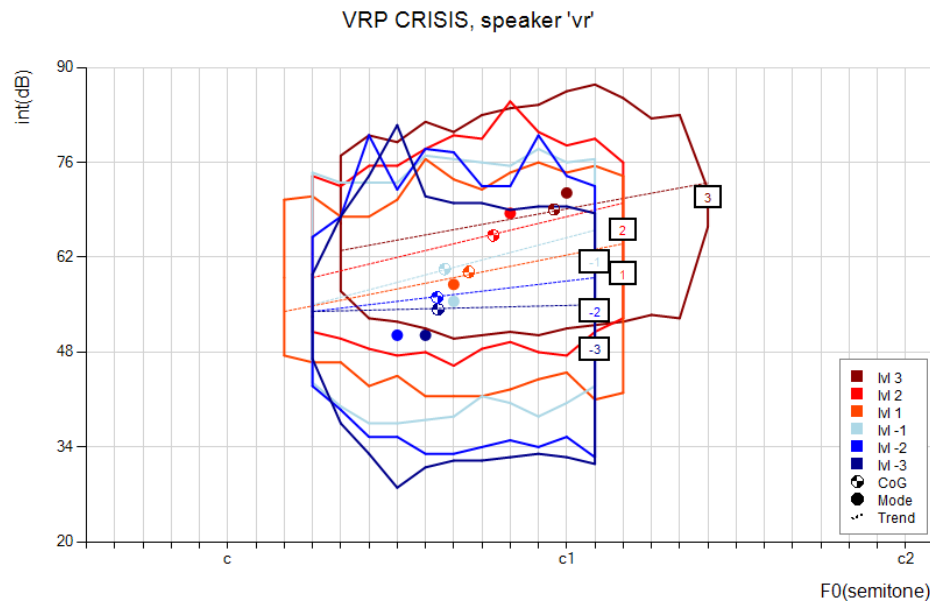
Stress cues in speech – Voice field (F0 versus Intensity)

Voice fields used for six levels of arousal by speaker MR with fully blown expression (a) and speaker VR with limited expressiveness (b). (Moreover MR is male and VR is female)

a)



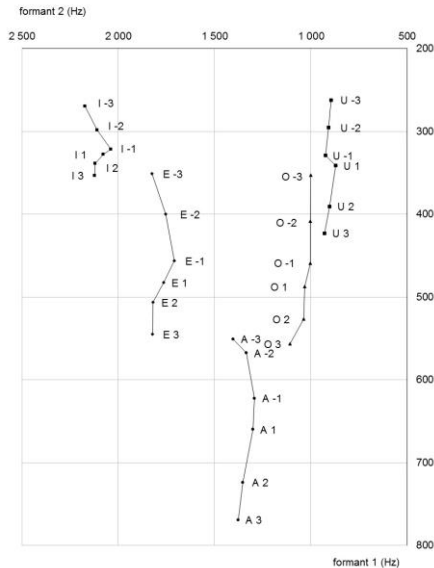
b)



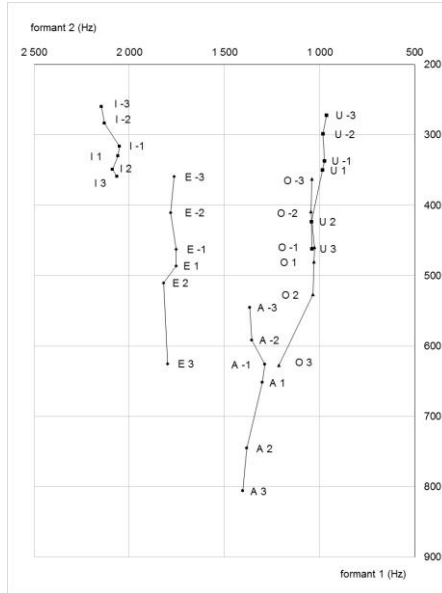
Stress cues in speech – Formants and spectra

First versus second (a) and second versus third formants (b) in a,e,i,o,u, vowels, and difference of LTAS spectra with respect to neutral speech LTAS spectrum (speaker MR, six levels of arousal)

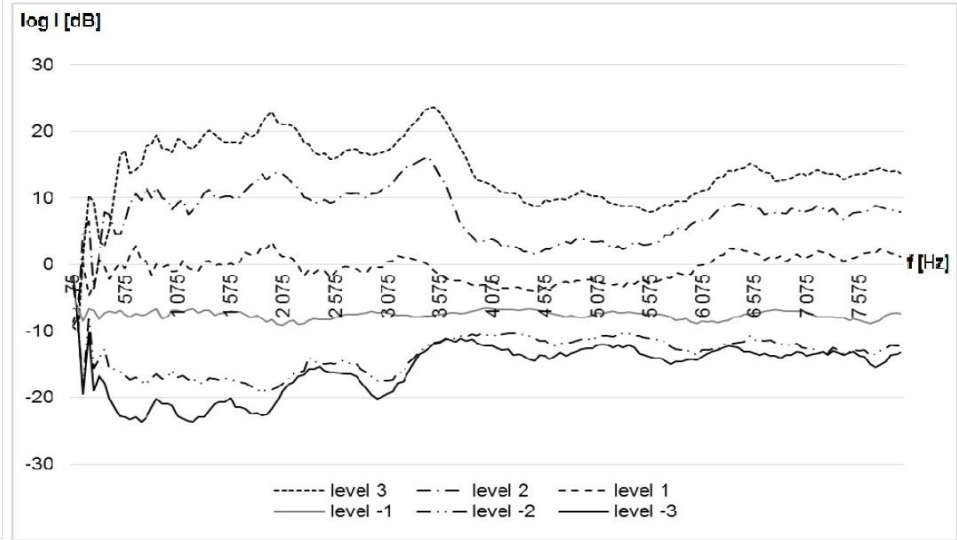
a)



(b)



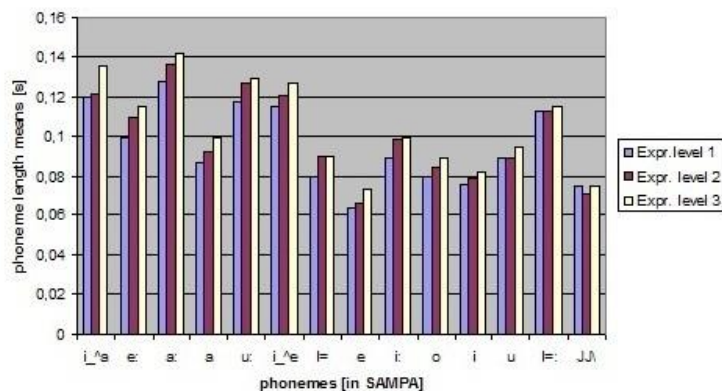
(c)



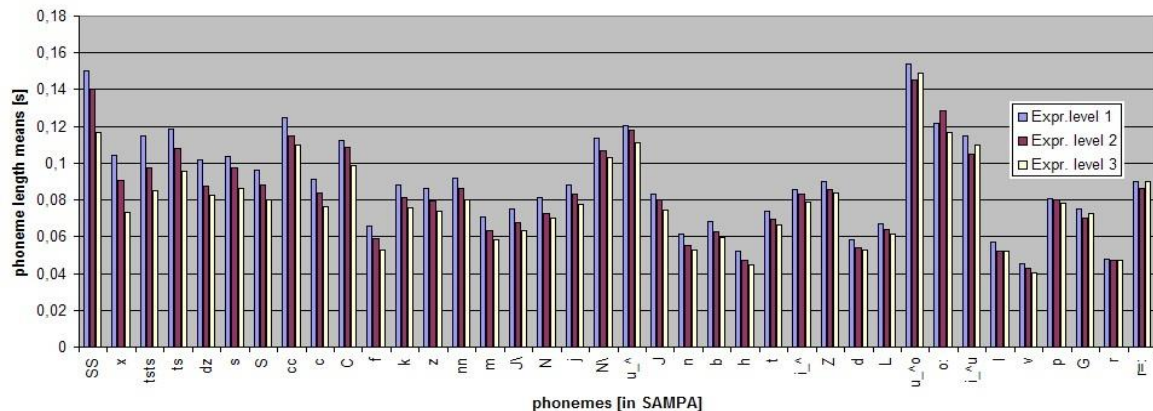
Stress cues in speech – Rhythm (segmental lengths)

Phonemes being lengthened (a) and phonemes being shortened (b) with increasing arousal (speaker MR, three levels of arousal)

a)

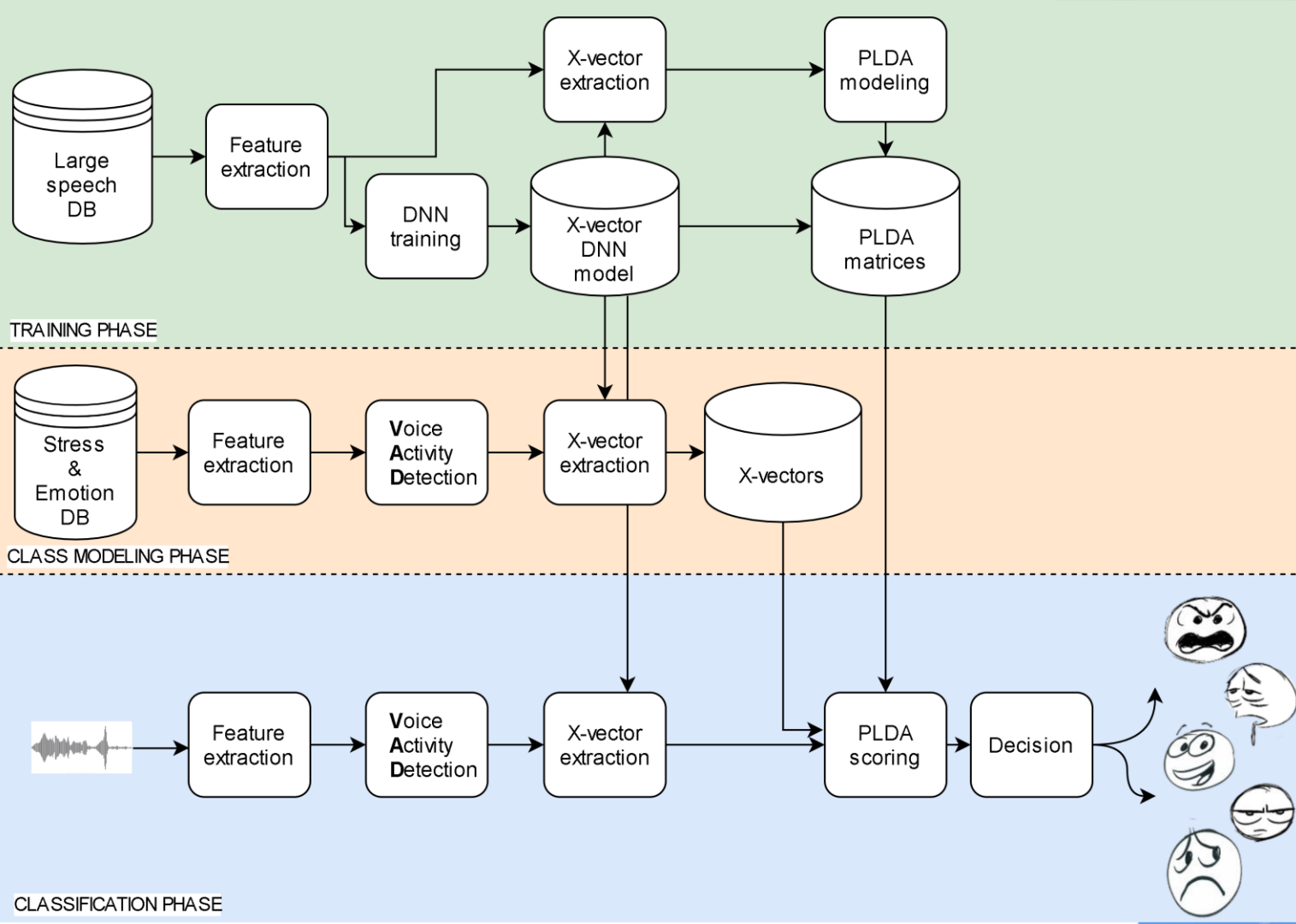


b)



Wovels and diphthongs, the segments carrying the highest load of energy are being lengthened, and sibilants, fricatives and plosives are being shortened.

Stress & Emotion recognition based on X-vectors



RESULTS: Leave-one-out cross-database-testing - Emotions

Full list of the emotional speech databases :

1. EMO-DB, 2. EmoV-DB_sorted, 3. EMOVO-Italian_Emotional_Database, 4. Enterface, 5. IEMOCAP ,
6. jl-corpus, 7. SAVEE, 8. RAVDESS, 9. Rockikz-emo, 10. MSP-Improv, 11. VESUS, 12. CREMA-D

7 classes of emotions:

1. *anger*
2. *disgust*
3. *fear*
4. *happy*
5. *neutral*
6. *sadness*
7. *surprise*

Accuracy in %

1. EMO-DB	14.98%
2. EmoV-DB_sorted	14.81%
3. IEMOCAP	32.31%
4. RAVDESS	32.13%
5. SAVEE	29.17%
6. CREMA-D	33.44%
7. EMOVO	32.82%
8. MSP-IMPROV	34.30%
(Chance, i.e. random choice	85.71%)

RESULTS: 4-class stress recognition based on arousal

In a 3-class test (low, neutral, high arousal) the recognizer tested on CRISIS database reached Error rate **7.78 %**

4 stress classes:

1. **Very low arousal**
level -0.5 to -1
2. **Neutral/normal**
arousal level around 0
3. **Increased** arousal
level around +0.5
4. **Highly increased**
arousal
level approaching +1

Confusion matrix for stress level recognition:

	Test low arousal	Test neutral arousal	Test high arousal	Test very-high arousal
Recognized low arousal	86.1%	16.1%	0.4%	0.0%
Recognized neutral	13.8%	73.3%	27.5%	4.0%
Recog. high arousal	0.0%	7.2%	55.7%	18.9%
Recog. very high arousal	0.1%	3.4%	16.3%	77.2%

Ongoing work and future steps

Ongoing research:

- Building and collecting emotional and stress databases, fusion into Emotional Data Pool
- Valence & Arousal recognition trained on categorically annotated emotional databases
- Channel mismatch (recognition on radio channel speech)

Future research:

- Optimization of features and recognizer architecture
- Dominance recognition (3D model)



Thank you for your attention

Announcement:

The work presented at this work is funded from the project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 832969. The outputs of the presented papers reflect the views only of the author(s), and the European Union cannot be held responsible for any use which may be made of the information contained therein. SATIE public website: <http://satie-h2020.eu/>

The presented work is also co-funded by European Regional Development Fund, Ministry of Transport and Construction of the Slovak Republic and Ministry of Economy of the Slovak Republic, in the frame of the project Early Warning of Alzheimer (reg. No. 67/2020-2060-2230-V631) and from the VEGA agency project No. 1/0667/18, Automatic assessment of acute stress from speech.