

# Rozpoznávání a syntéza řeči

PLIN059

Mgr. Dana Hlaváčková, Ph.D.

Mgr. Jakub Machura, Ph.D.

# Fonetika

- přírodní věda na pomezí lingvistiky, anatomie, fyziologie a fyziky (akustiky)
- zkoumá zvukovou stránku jazyky z různých aspektů
- Dělení fonetiky:
  - artikulační – tvorba fónů ve zvukovém ústrojí
  - akustická – přenos zvuků prostředím, jejich frekvence...
  - percepční – jak jsou zvuky přijímány

# Fonetika

- fón
- foném (např. lis – les – los).
- alofón
- koartikulace
- fonetická abeceda
- IPA

# THE INTERNATIONAL PHONETIC ALPHABET (revised to 2020)

## CONSONANTS (PULMONIC)

© 2020 IPA

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			ʀ					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

## CONSONANTS (NON-PULMONIC)

Clicks	Voiced implosives	Ejectives
◌ ɸ Bilabial	ɓ Bilabial	ʼ Examples:
◌ ǀ Dental	ɗ Dental/alveolar	pʼ Bilabial
◌ ǃ (Postalveolar)	ɟ Palatal	tʼ Dental/alveolar
◌ ǂ Palatoalveolar	ɡ Velar	kʼ Velar
◌ ǁ Alveolar lateral	ɠ Uvular	sʼ Alveolar fricative

## OTHER SYMBOLS

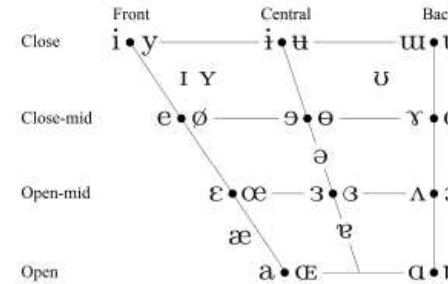
ɱ Voiceless labial-velar fricative	ç ʝ Alveolo-palatal fricatives
ɰ Voiced labial-velar approximant	ɻ Voiced alveolar lateral flap
ɟ Voiced labial-palatal approximant	ɥ Simultaneous ʃ and x
ħ Voiceless epiglottal fricative	Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary.
ʕ Voiced epiglottal fricative	
ʔ Epiglottal plosive	

## DIACRITICS

◌ ◌̥ Voiceless	◌ ◌̤ Breathy voiced	◌ ◌̦ Dental	◌ ◌̧ Apical
◌ ◌̣ Voiced	◌ ◌̥ Creaky voiced	◌ ◌̨ Laminal	◌ ◌̩ Nasalized
◌ ◌̪ Aspirated	◌ ◌̫ Linguolabial	◌ ◌̮ Labialized	◌ ◌̯ Palatalized
◌ ◌̬ More rounded	◌ ◌̭ Labialized	◌ ◌̮ Nasalized	◌ ◌̯ Nasal release
◌ ◌̮ Less rounded	◌ ◌̯ Palatalized	◌ ◌̮ Nasalized	◌ ◌̯ Nasal release
◌ ◌̯ Advanced	◌ ◌̮ Velarized	◌ ◌̮ Lateral release	◌ ◌̯ Lateral release
◌ ◌̮ Retracted	◌ ◌̯ Pharyngealized	◌ ◌̮ No audible release	◌ ◌̯ No audible release
◌ ◌̯ Centralized	◌ ◌̮ Velarized or pharyngealized	◌ ◌̮	
◌ ◌̮ Mid-centralized	◌ ◌̯ Raised	◌ ◌̮ Lowered	
◌ ◌̯ Syllabic	◌ ◌̮ Advanced Tongue Root	◌ ◌̯ Retracted Tongue Root	
◌ ◌̮ Non-syllabic	◌ ◌̯ Advanced Tongue Root	◌ ◌̮ Retracted Tongue Root	
◌ ◌̯ Rhoticity	◌ ◌̮ Advanced Tongue Root	◌ ◌̯ Retracted Tongue Root	

Some diacritics may be placed above a symbol with a descender, e.g. ɲ̥̄

## VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

## SUPRASEGMENTALS

ˈ Primary stress	ˈfoʊnəˈtɪʃən
ˌ Secondary stress	
ː Long	eː
ˑ Half-long	eˑ
◌̥ Extra-short	ẽ
◌̥ Minor (foot) group	
◌̥ Major (intonation) group	
◌̥ Syllable break	ai.ækt
◌̥ Linking (absence of a break)	

## TONES AND WORD ACCENTS

LEVEL	CONTOUR
ě or ǃ Extra high	ě or ǃ Rising
é ǂ High	é ǂ Falling
ē ǁ Mid	ē ǁ High rising
è ǁ Low	è ǁ Low rising
ë ǁ Extra low	ë ǁ Rising-falling
↓ Downstep	↘ Global rise
↑ Upstep	↙ Global fall

[https://en.wikipedia.org/wiki/International\\_Phonetic\\_Alphabet](https://en.wikipedia.org/wiki/International_Phonetic_Alphabet)

# Fonetický přepis

*Děti se zítra čestně přiznají, že okno rozbily.*

přepis IPA:

[ ɟɛci sɛ zɪːtra tʃɛstɲɛ pɹɪznajɪː | zɛ ʔokno rozbɪlɪ || ]

český přepis:

[ d'etɪ se zɪːtra ɟɛstɲɛ pɹɪznajɪː | žɛ ʔokno rozbɪlɪ || ]

# Rozpoznávání řeči

- Automatic **Speech Recognition** (ASR)
- Speech-To-Text (STT)
- odlišit od – **rozpoznávání hlasu** – identifikace mluvčího (verifikace, bezpečnostní systémy)
- **převod zvukového signálu na text**
  - bigramy, n-gramy, pravděpodobnost výskytu slov na základě předchozího kontextu, neuronové sítě
  - velikost slovníku, korpusy
- vliv intonace, výslovnosti, přízvuku a okolního šumu
- komplikovanější pro flektivní jazyky

# Rozpoznávání řeči

- rozpoznávání **izolovaných slov**
  - hlasové povely – hlasové ovládání počítače, hry, automobilu
- rozpoznávání **spojité řeči**
- jeden mluvčí – Speaker Dependent
- více mluvčích – Speaker Independent (možnost adaptace)
- **meze:**
  - mikrofon + eliminace šumu
  - zřetelná výslovnost
  - doménově specializovaná oblast
    - omezená slovní zásoba (právo, medicína)
  - natrénováno na jednoho mluvčího

# Rozpoznávání řeči

- ověřování hlasem
- diktovací systémy
- transkripce audio nahrávek
- přepisy záznamu televizních a rozhlasových pořadů
- titulkovací systémy
- pomoc handicapovaným osobám (hlasové ovládání)
- dialogové systémy, automatická spojovatelka
- mobilní aplikace



# Rozpoznávání řeči

- Transkriptor – <https://transkriptor.com/cs/>
- [Funkce Transcribe od Microsoft](#)

# Pracoviště v ČR

- ZČU Plzeň – **Katedra kybernetiky** Fakulty aplikovaných věd, Oddělení umělé inteligence, <http://www.kky.zcu.cz/cs>
  - akustické rozpoznávání řeči
  - audiovizuální rozpoznávání řeči
  - rozpoznávání znakové řeči
  - titulkovací systémy
- původně spin-off **SpeechTech** (<https://www.speechtech.cz>)
  - diktovací systém NovaVoice (technologie MegaWord), se společností Consulting Company Novasoft (<https://ccnovasoft.cz>)

# Pracoviště v ČR

- TU Liberec – Ústav informačních technologií a elektroniky  
Fakulty mechatroniky, informatiky a mezioborových studií
- **Laboratoř počítačového zpracování řeči SpeechLab**  
(<http://speechlab.tul.cz/index.html>)
  - diktovací systém NewtonDictate, s firmou Newton Technologies (<https://www.newtontech.net/cs/>)
- VUT Brno – Fakulta informačních technologií
  - **Ústav počítačové grafiky a multimédií**
  - **Speech@FIT** – výzkum zaměřený na verifikaci hlasu
  - původně spin-off Phonexia (<http://www.phonexia.com/>)

# Ukázky

Katedra kybernetiky

<https://www.youtube.com/channel/UCgLPJ6VT9rmwQ3GuZBajZ2w>

NewtonDictate

- <http://www.youtube.com/watch?v=W0BT6DKzF3s>
- a další videa na YouTube...

pro pobavení:

- Windows Vista Speech Recognition  
<http://www.youtube.com/watch?v=kX8oYoYy2Gc>

# Syntéza řeči

- **Speech Synthesis**
- Text-To-Speech (TTS)
- subslovní řečové segmenty, konkatenace (řetězení)
- **difóny** (jednotka od poloviny jedné hlásky do poloviny druhé)
- **trifóny** (kontext hlásky)
- modelování prozodických charakteristik řeči (melodie, tempo, hlasitost)
- mluvené (řečové) korpusy
- kvalita posuzována podle podobnosti lidskému hlasu

# Syntéza řeči

- služby telefonních systémů
- software pro zrakově postižené osoby – odečítače obrazovky (screenreader, *pristupnost.cz/screen-readery*)
- ZČU <http://www.kky.zcu.cz/cs>
- spojení s vizualizací (audiovizuální syntéza řeči)
- SpeechTech <https://www.speechtech.cz>
- Acapela (<https://www.acapela-group.com/>)
- Google WaveNet – neuronové sítě, <https://cloud.google.com/text-to-speech>

# Dialogové systémy

- Google Assistant <https://assistant.google.com/>
- Amazon Alexa [https://cs.wikipedia.org/wiki/Amazon\\_Alexa](https://cs.wikipedia.org/wiki/Amazon_Alexa)
- ELIZA computer therapist <https://cs.wikipedia.org/wiki/ELIZA>
- Talk to Poppy / Elysai – terapie <https://www.talktopoppy.com/>
- <https://promethist.ai/>
- iPhone – asistentka Siri, rozumí a odpovídá  
[https://www.youtube.com/watch?v=AfzPngLC92U&ab\\_channel=MayankBhardwaj](https://www.youtube.com/watch?v=AfzPngLC92U&ab_channel=MayankBhardwaj)
- [SMART HOUSE – ENGLISH VERSION – REMA 1000 - YouTube](#)