

## Umí váš telefon mluvit?

**Jak zkoumat zvukovou stránku jazyků a jak tyto poznatky použít mimo jazykovědu?**

- tvorba a vnímání řeči
- mluvící stroje a poslouchající stroje
- fonetika jako nástroj pro odhalení zločinu

**Aleš Bičan**

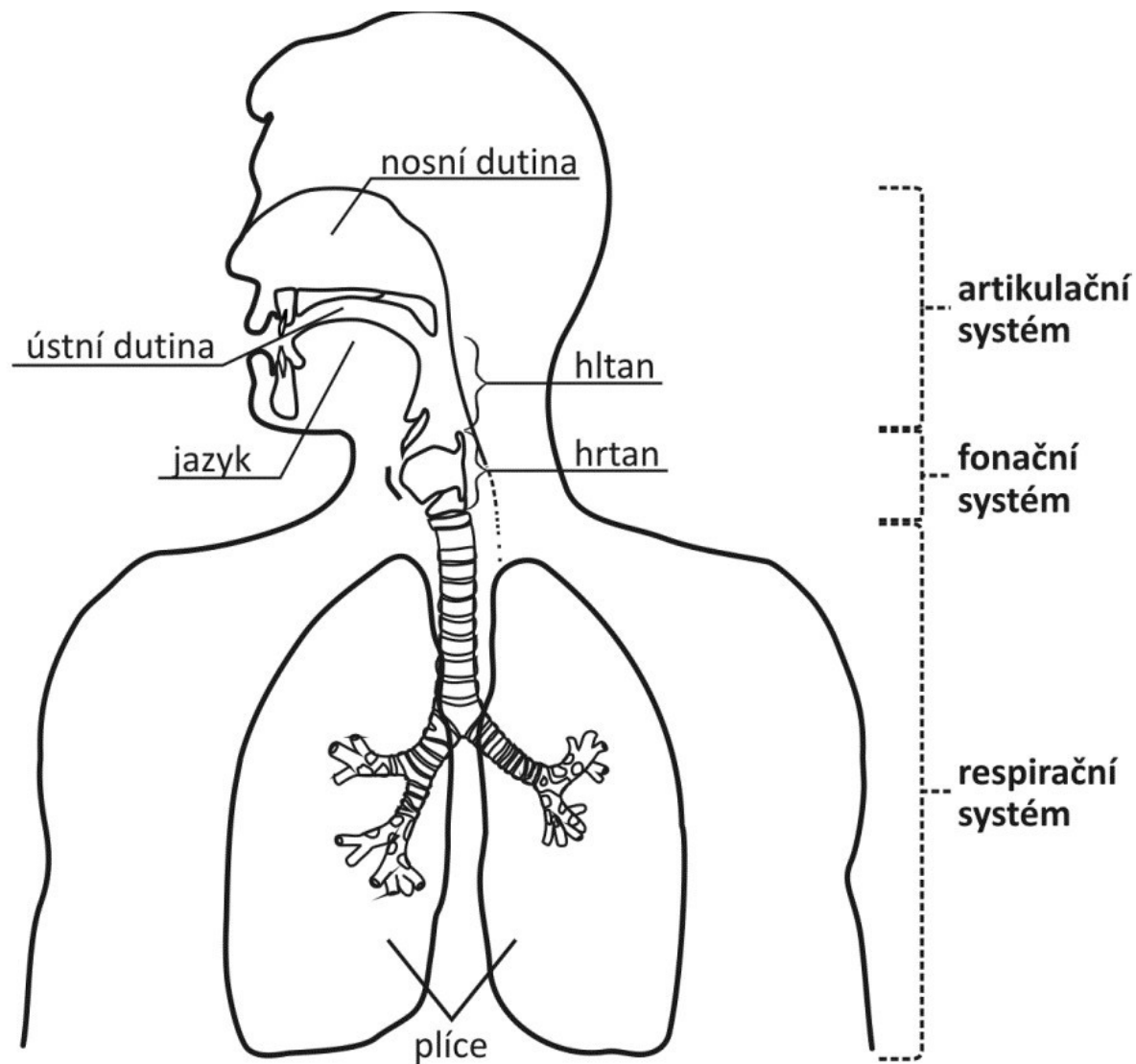
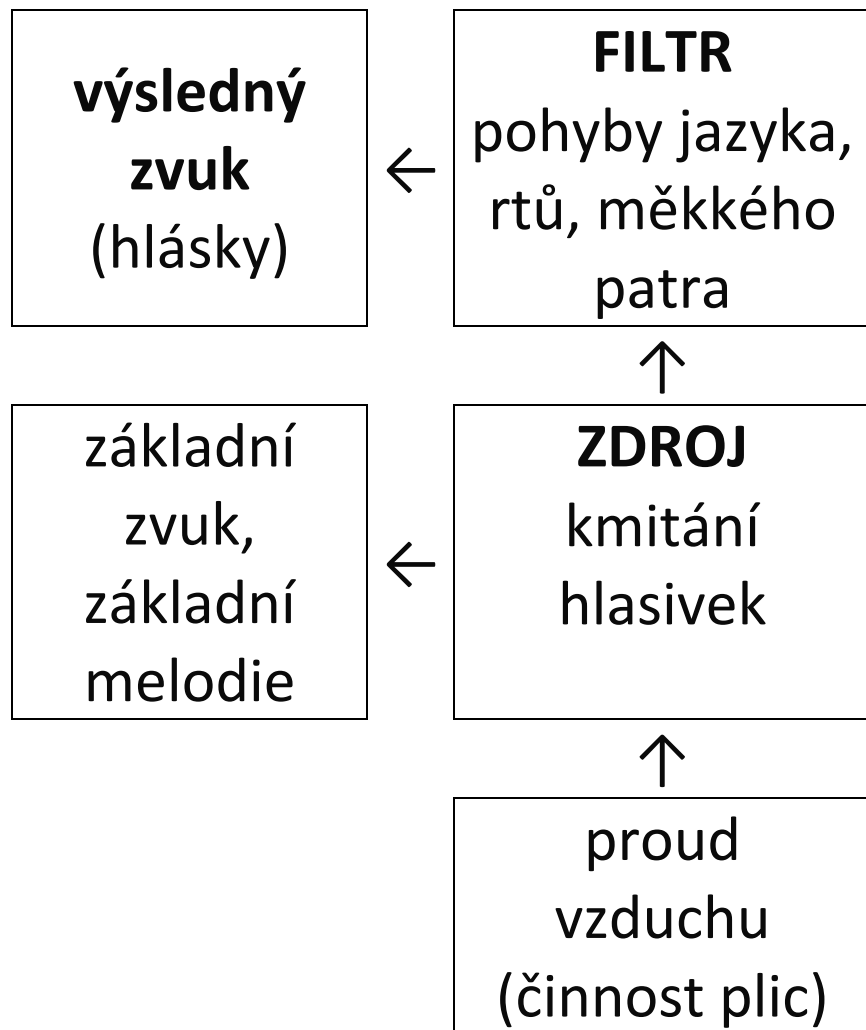


## Jaké aspekty řeči chceme zkoumat a jaké nástroje pro to použijeme?

| <i>Co zkoumat?</i>      | <b>tvorba řeči</b>   | <b>vnímání a porozumění řeči</b>   | <b>řeč jako druh zvuku</b>  | <b>využití řeči pro komunikaci</b>   |
|-------------------------|--|--|---|--|
| <i>Jak zkoumat?</i>     | <ul style="list-style-type: none"> <li>• přímé pozorování</li> <li>• rentgen</li> <li>• umělé patro</li> </ul> | <ul style="list-style-type: none"> <li>• přímé pozorování</li> <li>• nahrávky</li> </ul> | <ul style="list-style-type: none"> <li>• přístroje analyzující zvuk</li> <li>• počítačové programy<sup>[*]</sup></li> </ul> | <ul style="list-style-type: none"> <li>• přímé pozorování</li> <li>• chování mluvčích</li> <li>• analýza jazykových dokladů</li> </ul> |
| <i>Jaká disciplína?</i> | <b>ARTIKULAČNÍ</b>   | <b>PERCEPČNÍ</b>   | <b>AKUSTICKÁ</b>  | <b>FONOLOGIE</b><br>(„funkční fonetika“)   |
|                         | <b>FONETIKA</b>  |  |   |  |

[\*] např. Praat, <https://www.fon.hum.uva.nl/praat/>

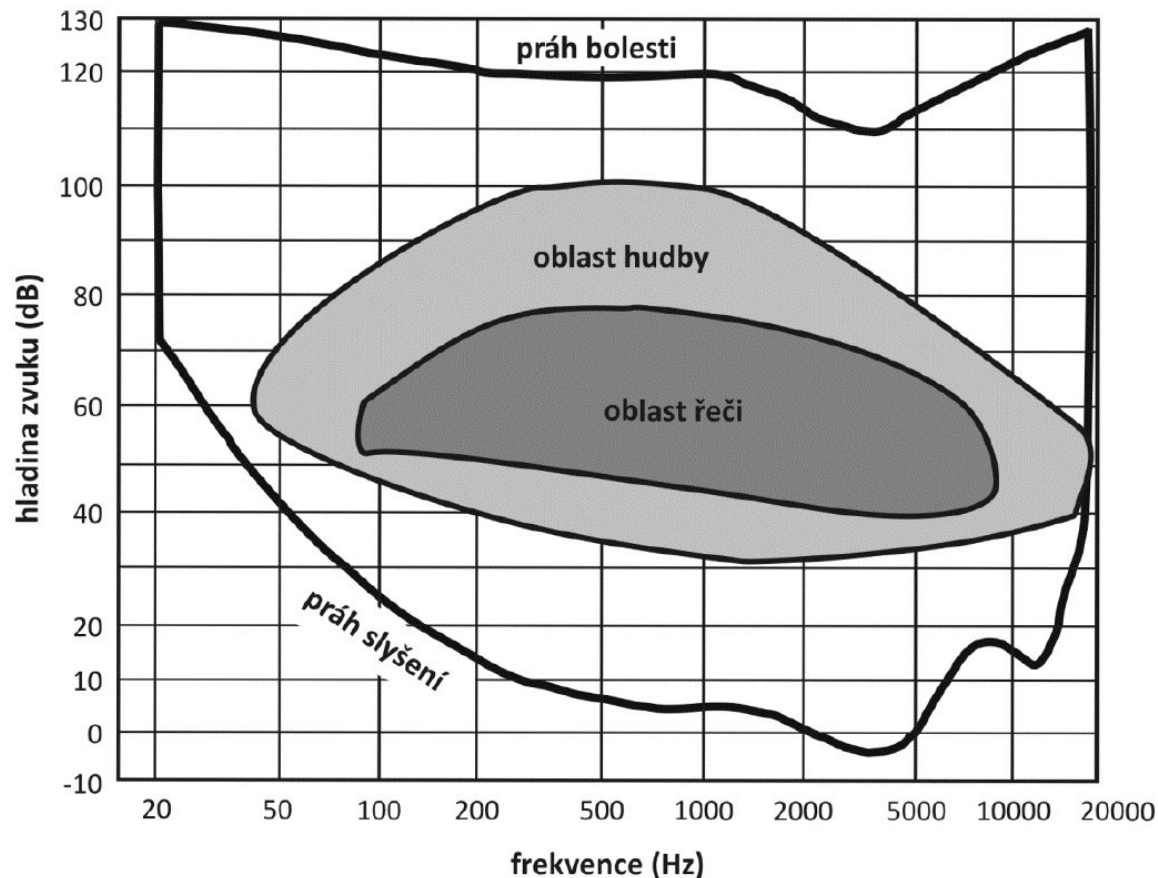
## Tvorba řeči: artikulační fonetika



## Vnímání a porozumění řeči: *percepční fonetika*

- Lidské ucho je schopno vnímat jen určitý rozsah zvuků (**biologický filtr**)
- Vnímání zvuků je však **selektivní**: malé dítě je schopno rozpoznat všechny zvuky lidských jazyků, ale postupně tuto schopnost ztrácí a omezuje se pouze na zvuky svého jazyka (např. Japonci [r] *right* × [l] *light*, Češi [n] *sin* × [ŋ] *sing*)
- Posloucháme rty a vidíme zvuky aneb nevnímáme pouze sluchem, ale i **zrakem**<sup>[\*]</sup>

[\*] McGurkův efekt, <https://www.youtube.com/watch?v=PWGeUztTkRA>



- Vnímání může být ovlivněno tzv. **primingem** (priming): dřívější podnět ovlivňuje pozdější chování a myšlení

➤ *Případ Davida Baina*

(odsouzen za vraždu 1995, později osvobozen v odvolacím řízení)

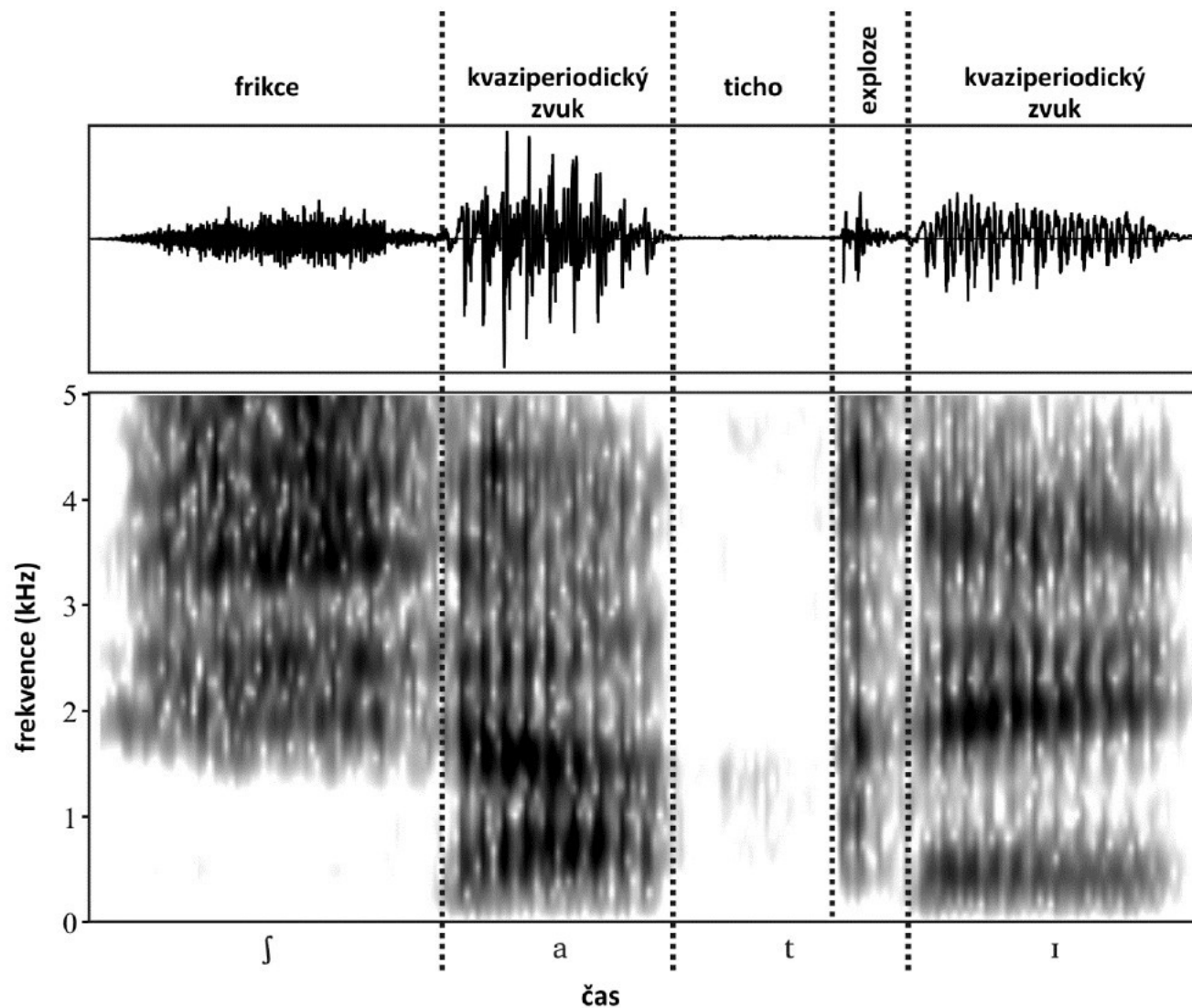
- Řekl Bain v telefonu *I shot the prick.*?



## Řeč jako zvuk: *akustická fonetika*

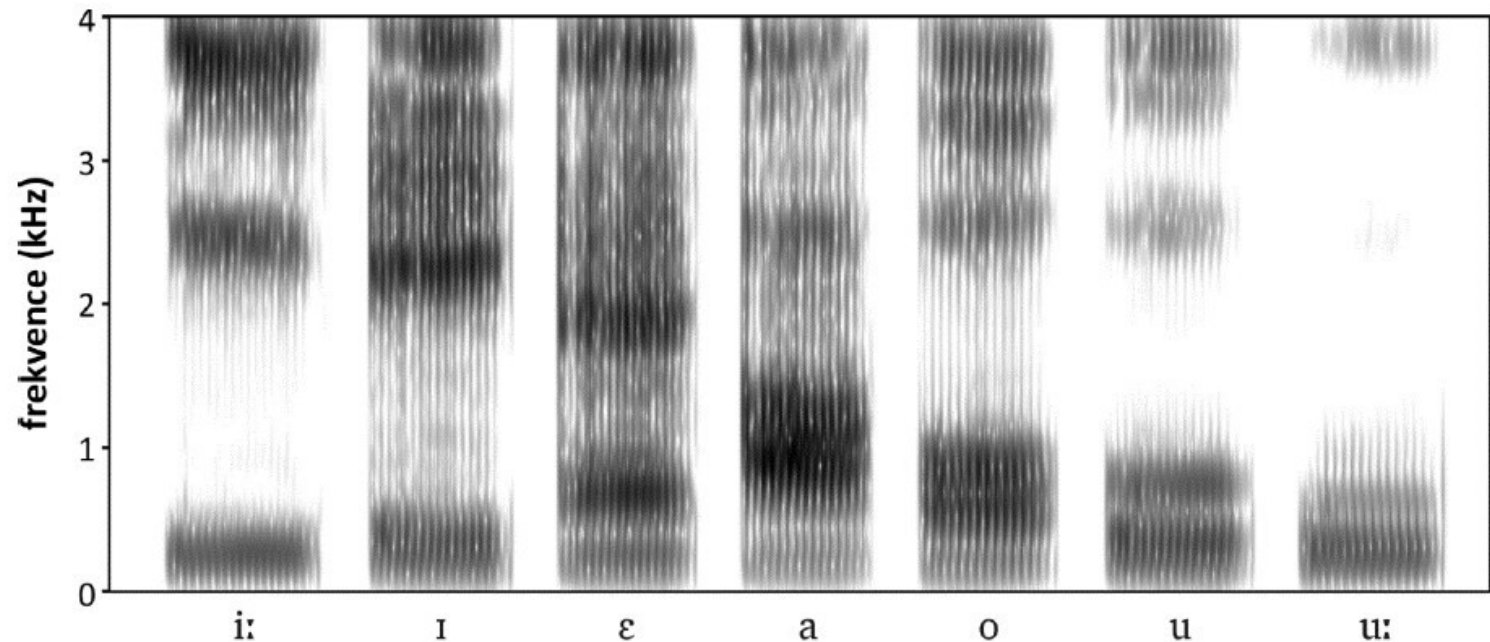
Řeč je kombinace

- **tónových zvuků**  
(pravidelné kmitání a rezonance)
  - **šumových zvuků**  
(nepravidelné šumy vzniklé třením)
  - **ticha** (při uzavření průchodu vzduchu)
- 
- Tónové zvuky lze rozdělit na různé složky, každá má jinou **frekvenci** a **intenzitu**
  - Nejvýraznější jsou samohlásky díky svým **formantům** (frekvenční složky s výraznou intenzitou – tmavé pruhy ve **spektrogramu**)



## Existuje unikátní otisk hlasu stejně jako unikátní otisk prstu?

- Nejspíš ne, protože hlas jedince je odrazem jeho **fyziognomie** (velikost hlavy a dutin, velikost a tvar hlasivek) a **výchovy** (naučíme se mluvit určitým způsobem) [*nature* × *nurture* : dědičnost × prostředí]
- Každý má jiný **idiolekt** projevující se jako specifická výslovnost (**akcent**), jako odraz regionální variability (**dialekt**) nebo sociální variability (**sociolekt**)
- I jednovaječná identická dvojčata mají jiný hlas



## To je všechno pěkné, ale k čemu to je?

- Jak poznatky o lidské řeči použít pro zjednodušení a zlepšení lidského života?
- Jak mohou tyto poznatky napravit, co jsme my nebo příroda „pokazila“?
- Jak může být fonetika užitečná v různých odvětvích lidské činnosti (pracovní, umělecké, právní, vědecké atd.)?

### Fonetika

- Jazykovědná disciplína s největším přesahem k přírodním vědám
- Jazykovědná disciplína s největší využitelností v praxi
- Mediálně zajímavá

např. Radek Skarnitzl (ředitel Fonetického ústavu FF UK) v *Show Jana Krause*:

<https://www.youtube.com/watch?v=ZjKbBB-JpmQ>



## Možnosti aplikace fonetiky v praxi (výběr)

|  |  |   |
|--|--|---|
| <b>Informační technologie, strojírenství (řečové technologie)</b>  | <b>Kriminalistika (forenzní fonetika)</b>  | <b>Lékařství (foniatrie, logopedie)</b>   |
| <ul style="list-style-type: none"><li>• vytvoření umělé řeči</li><li>• automatické rozpoznání řeči</li><li>• automatické rozpoznání jazyka</li></ul> | <ul style="list-style-type: none"><li>• rozpoznání mluvčího</li><li>• identifikace mluvčího</li><li>• profilování mluvčího</li><li>• rozpoznání toho, co bylo řečeno</li></ul> | <ul style="list-style-type: none"><li>• odstranění vrozených a poúrazových vad, poruch řeči</li><li>• nápravy řečových vad a chybné výslovnosti</li></ul> |

|   |  |  |
|---|--|--|
| <b>Vzdělávání</b>   | <b>Umění</b>   | <b>A další...</b>  |
| <ul style="list-style-type: none"><li>• výuka cizích jazyků</li><li>• výuka správné výslovnosti</li></ul> | <ul style="list-style-type: none"><li>• tvorba jazyků</li><li>• řeč jako herecký prostředek</li><li>• zpěv</li></ul> | <ul style="list-style-type: none"><li>• např. reklama, tvorba značek (branding), animace</li></ul> |

## Mluvicí stroje: syntéza řeči

- Může lidskou řeč vytvořit něco jiného než člověk?
- Lze vytvořit mechanický mluvicí stroj?[\*]
- Lze elektronicky generovat řeč stejně jako jiné zvuky?

[\*] Wolfgang von Kempelen (1734–1804),

[https://www.youtube.com/watch?v=k\\_YUB\\_S6Gpo](https://www.youtube.com/watch?v=k_YUB_S6Gpo)



Praktické výhody:

- Lidé s poruchou řeči by mohli mluvit
- Slepí lidé by mohli mít přístup k psanému textu
- Líní lidé by nemuseli číst, ale jen poslouchat
- Přístroje by mohly nahradit práci lidí (při poskytování informací přes telefon, automatický dabing)

*Alexa (Amazon), Siri (Apple), Cortana (Microsoft)*

# Jak probíhá syntéza řeči?

- Převod textu do řeči (**TTS**: text to speech)
- Převod není přímý, ale probíhá přes **fonetickou transkripci**

[ | | 'fonetikaʊje | 'jazikovjedna: | 'distɕipli:na | 'zabi:vaji:tsi:ʊse | 'materija:lɲi: | 'stra:ŋkou | 'zvukovi:x | 'vi:razovi:x | 'prostřetku: | 'jazika | | ]

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2005)

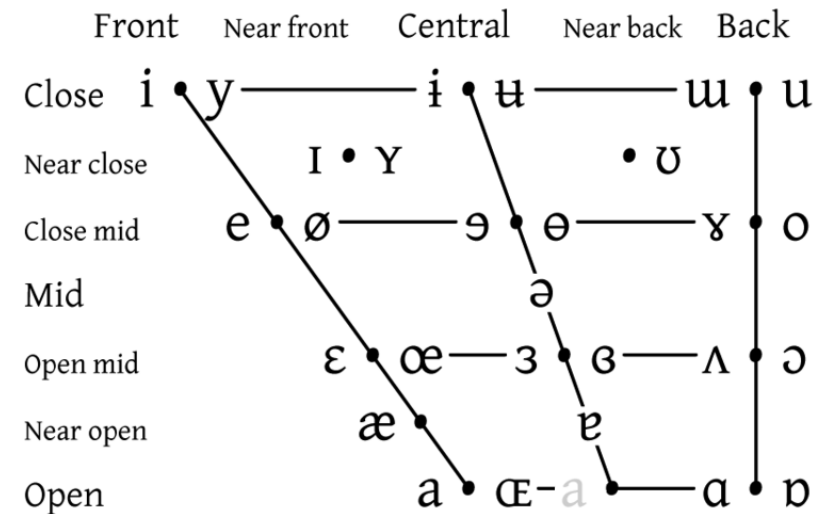
CONSONANTS (PULMONIC)

© 2005 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        |             |        | ɾ        |              |           |         |       | ʀ      |            |         |
| Tap or Flap         |          | ⱱ           |        | ɽ        |              | 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     |        |            |         |

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

## VOWELS



Vowels at right & left of bullets are rounded & unrounded.

## Psaný jazyk ≠ mluvený jazyk

- Zapisujeme i to, co nevyslovujeme (řeč se rychle mění, psaný jazyk méně)  
čeština: *jsi*, *leccos*, *srdce*  
angličtina: *late*, *knight*, *write*  
mezery (!): *dá mi* × *dámy*
- Vyslovujeme i to, co nepíšeme (více, než si myslíme)  
přízvuk: angl. *import*<sub>podš. jm.</sub> × *import*<sub>sloveso</sub>, č. *na stole* × *stole*  
některé hlásky: [fʔoku] *v oku*, [marije] *Marie*
- Pravopis je víceznačný (jeden **grafém** = více **fonů**)  
*pouček* [pouček] × [poʔuček]  
*ion* [jo:n] × *špion* [špijo:n]  
*baby* [babi] × [bejbi]

|                           |                             |                                |                                 |                                |                             |                              |                            |
|---------------------------|-----------------------------|--------------------------------|---------------------------------|--------------------------------|-----------------------------|------------------------------|----------------------------|
| like <b>o</b><br>in<br>go | like <b>oo</b><br>in<br>too | like <b>off</b><br>in<br>offer | like <b>uff</b><br>in<br>suffer | like <b>ow</b><br>in<br>flower | like <b>aw</b><br>in<br>saw | like <b>a</b><br>in<br>above | like <b>up</b><br>in<br>up |
| though                    | through                     | cough                          | rough                           | plough                         | bought                      | borough                      | hiccough                   |
| although                  |                             | trough                         | tough                           | drought                        | fought                      | thorough                     |                            |
| dough                     |                             |                                | enough                          |                                | thought                     | thoroughfare                 |                            |
| doughnut                  |                             |                                |                                 |                                | ought                       |                              |                            |

- Někdy zapisujeme celé pojmy  
100 % studentů × 100% úspěšnost  
777 123 567
- Jindy zase zkracujeme  
sv. Tomáš × sv. první  
IPA [ipa] × [i:pe:a:]

## Vlastní syntéza: generování umělé řeči

|                    | Přístupy k syntéze řeči  |   |
|--------------------|--|---|
|                    | parametrický   | konkatenační  |
| Východiska:        | <ul style="list-style-type: none"><li>● akustické vlastnosti řeči zaznamenané jako číselné hodnoty</li></ul>   | <ul style="list-style-type: none"><li>● reálné promluvy (které se rozloží, složí a smíchají)</li></ul>  |
| Metoda:            | <ul style="list-style-type: none"><li>● zřetězení modelů řečových jednotek pomocí statistických modelů</li></ul>   | <ul style="list-style-type: none"><li>● zřetězení vhodně vybraných reálných zvukových jednotek</li></ul>  |
| Výhody a nevýhody: | <ul style="list-style-type: none"><li>● stabilní kvalita, nižší kvalita řeči kvůli zprůměrování hodnot parametrů</li><li>● vysoká adaptabilita</li></ul> | <ul style="list-style-type: none"><li>● větší paměťová náročnost, ale přirozenější řeč</li><li>● občasný výrazný pokles kvality u neobvyklých spojení řečových jednotek</li></ul> |

Ukázky obou:

<http://www.kky.zcu.cz/cs/research-fields/acoustic-speech-synthesis>

## Co se skládá při (konkatenací) syntéze?

**Slova**, to je jasné!

- Časově náročné

velké množství slov a slovních tvarů

asi 120 000 podstatných jmen × 7 pádů × 2 čísla = 1 680 000 různých tvarů podstatných jmen v češtině

- Náročné na hardware a software

nutnost uložit nahrávky všech slov a rychle mezi nimi vyhledávat

- Prakticky nemožné

žádná databáze nezaznamenává všechna slova (často chybí **nářeční slova, vlastní jména**)

nová slova stále přibývají (**neologismy, přejímky**)

## Tak **hlásky!**

- Navzdory hláskovému písmu po hláskách vůbec nemluvíme  
*pes* není [p] [e] [s]  
malé děti a negramotní lidé nejsou obvykle schopni hlásky rozeznat
- Řeč je nepřetržitá, hlásky na sebe plynule navazují a navzájem se ovlivňují  
důsledkem **koartikulace**  
[ki]tara × [ku]ře
- Reálnější jednotkou tvorby a vnímání řeči je **slabika** (ale podobné problémy)



## Tak co tedy? **Difony!**

- Difon (di-fon): spojení dvou vedle sebe stojících hlásek (fonů), popř. ticha  
*strana* = ticho-[s] + [st] + [tr] + [ra] + [an] + [na] + [a]-ticho

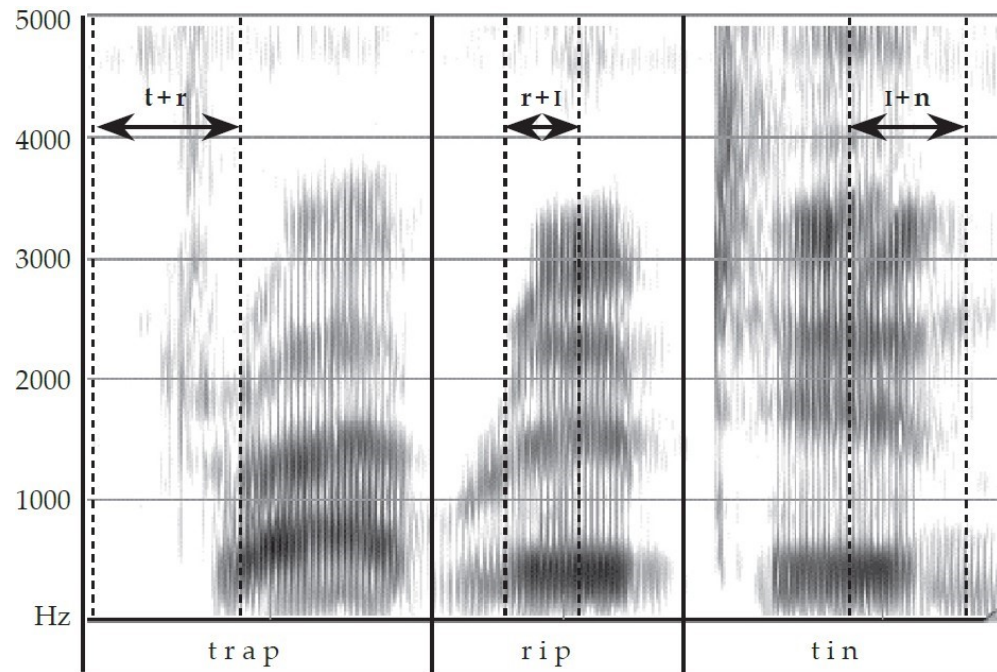
- „Slepí“ se dohromady a překryvy se upraví

- Dnes se však používá kombinace jednotek různé velikosti a volí se nejlepší kandidát

*stranám*: [stran] + [na:m]

*stranovat*: [stra]<sub>strana</sub> + [anovat]<sub>panovat</sub>

- I náš mozek nejspíš takto řeč ukládá a tvoří



## To ale nestačí!

- Je nutné vzít v úvahu rozdíly v melodii (**intonaci**) a síle řeči (**přízvuku**)

*import*<sub>podst.jm.</sub> × *import*<sub>sloveso</sub>

*pole* × *na pole*

*To je moje.* × *To je moje?* × *To je moje!*

*Upéct ne, uvařit.* × *Upéct, ne uvařit.* × *Upéct, neuvařit.*

## Jak dobrá je syntéza řeči?

- Dvě protichůdné síly: **srozumitelnost** × **přirozenost**
- Vysoce srozumitelné syntézy málo přirozené
- Vysoce přirozené syntézy mohou být méně srozumitelné
- V reálné řeči dochází k zjednodušování a splývání hlásek

*A já mu povídám, ty vole, to nemůžeš šidit, protože to on pozná!*

[ʔa ja: mu povi:da:m ti vole to nemu:žeš šidit protože to ʔon pozna:]

×

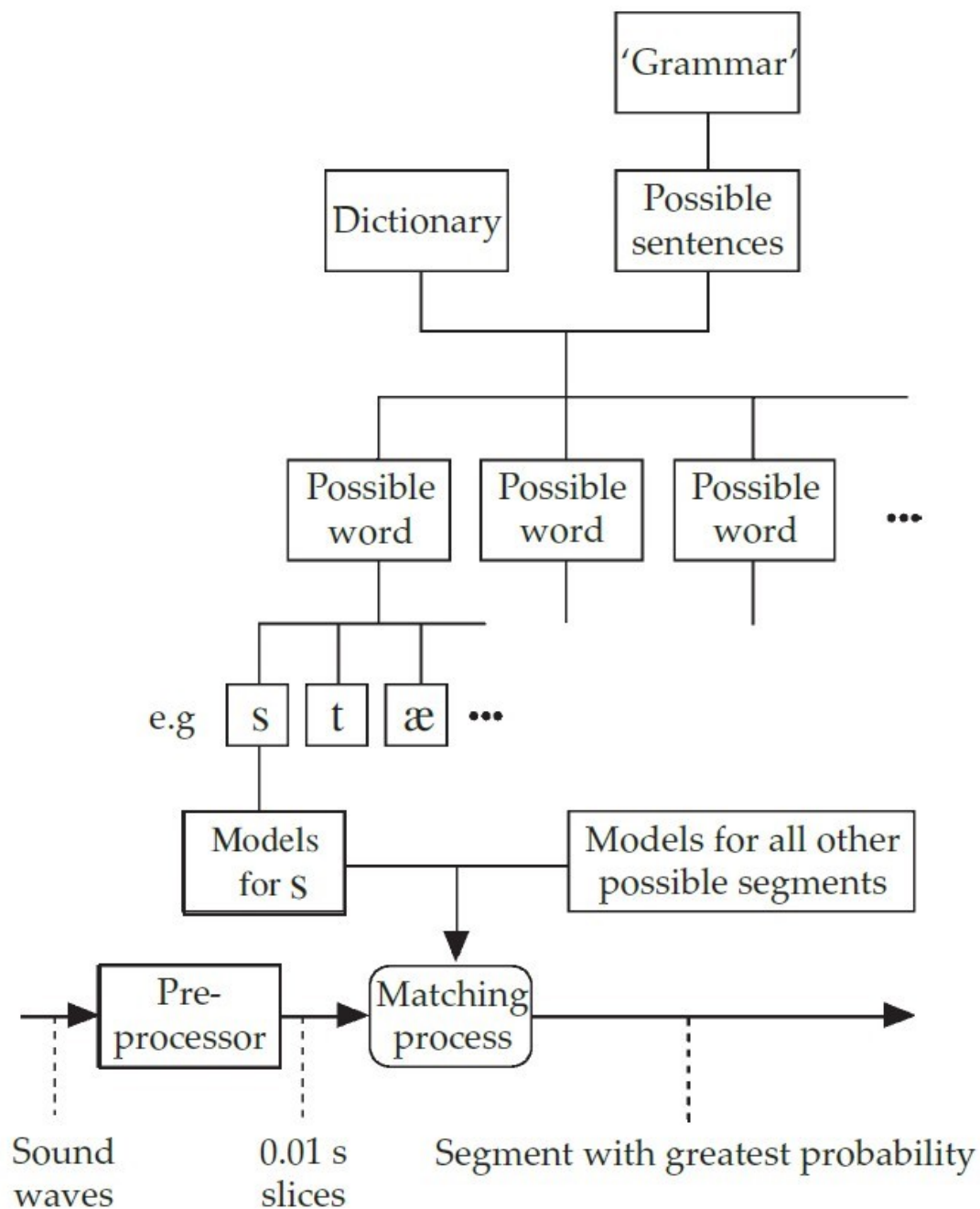
[ʔa ja m̩ pa:m ti we to nem̩že šidit p̩še to:n pozna:]

## Poslouchající stroje: Automatické rozpoznání řeči (ARŘ)

- Daleko složitější než syntéza řeči
  - u syntézy stačí vytvořit jeden hlas, ale u ARŘ pracujeme s miliony různých hlasů
- Proto se systémy obvykle omezují na konkrétní mluvní situace
  - např. komunikace s bankou přes telefon, informace o letech
- Nebo omezeny na konkrétního mluvčího, kterého se **učí** rozpoznávat

## Jak ARŘ probíhá?

- Nahrávka se rozdělí na úseky po 10 ms
- Počítač (neuronová síť) srovnává každý úsek s databází zvuků a stanovuje míru pravděpodobnosti, jakou zvuk odpovídá různým hláskám
- Zároveň se vypočítává kontextuální (**fonotaktická**) pravděpodobnost pro danou hlásku
- A také jazyková (**syntaktická**) pravděpodobnost pro dané slovo a posloupnost slov
- Výsledek je **nejpravděpodobnější** shoda



## Fonotaktika

- Kombinace hlásek (fonémů) nejsou v jazyce náhodné
- Některé kombinace nejsou možné
  - [t'] + [e:]
  - [t] + [č] na začátku slov
  - [ř] + [j] uvnitř kořene
- Některé kombinace jsou častější než jiné
  - na začátku slov: [př] > [pr] > [st] > [sp] > [kr] > [tr] > [sl]

Jak to zjistit?

- Rozborem zvukové struktury slov nebo experimenty s mluvčími
- Pro češtinu existuje *Fonologický korpus češtiny* (transkripce 461 tisíc slov)<sup>[\*]</sup>

<sup>[\*]</sup> <https://ujc.avcr.cz/phword>

## Jaká je v češtině pravděpodobnost dalších hlásek po nosovce?

### 1. hláska

### 2. hláska

### 3. hláska

|         |           |
|---------|-----------|
| nosovka | [m]: 32 % |
|         | [n]: 62 % |
|         | [ň]: 2 %  |

|                   |
|-------------------|
| samohláska: 82 %  |
| souhláska: 18 %   |
| samohláska: 100 % |
| souhláska: 0 %    |
| ...               |

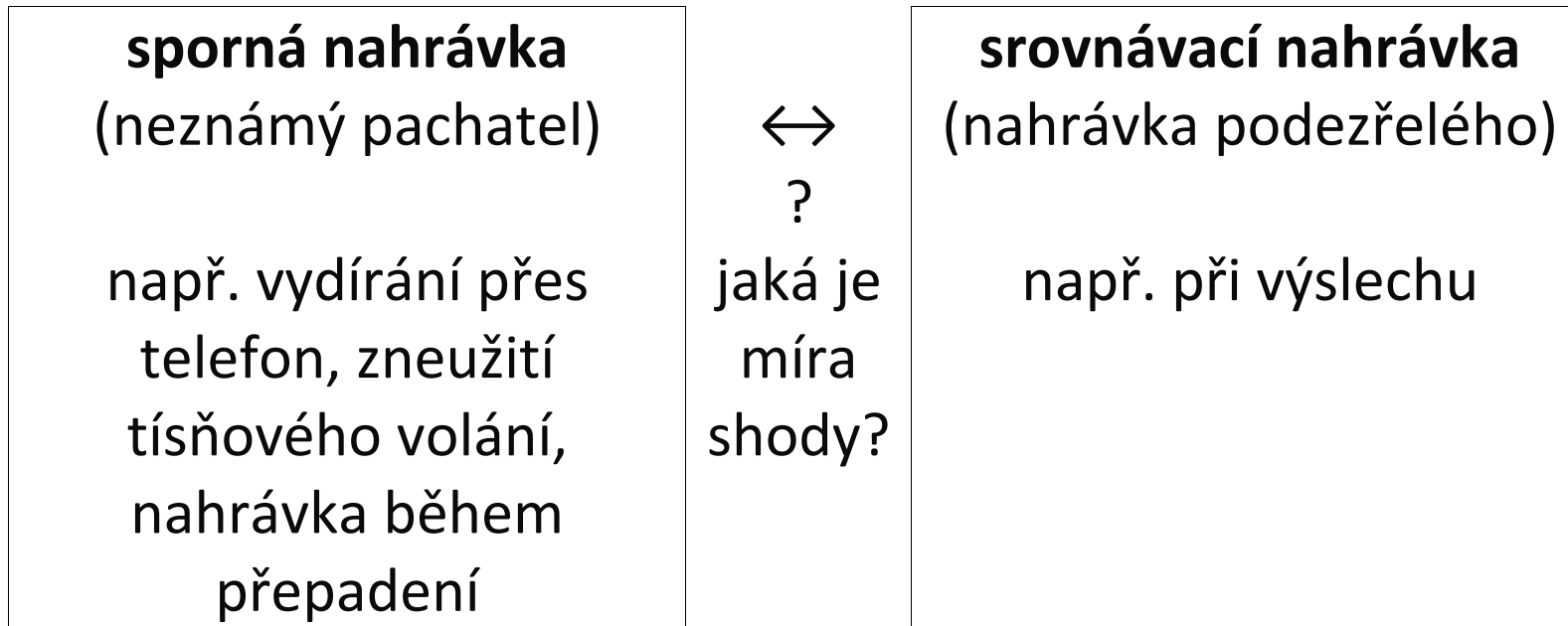
|                 |           |
|-----------------|-----------|
| krátká s.: 90 % | [a]: 35 % |
|                 | [e]: 17 % |
|                 | [i]: 21 % |
|                 | [o]: 20 % |
|                 | [u]: 7 %  |
| dlouhá s.: 10 % | ...       |
| ...             | ...       |
| krátká s.: 93 % | [a]: 30 % |
|                 | [e]: 62 % |
|                 | [i]: 1 %  |
|                 | [o]: 5 %  |
|                 | [u]: 1 %  |
| dlouhá s.: 7 %  | ...       |
| ...             | ...       |
| ...             | ...       |

## Forezní fonetika *aneb* čím může být fonetika prospěšná pro kriminalistickou praxi?

### *Některé úkoly forezní fonetiky*

| <b>Identifikace mluvčího</b>  | <b>Profilování mluvčího</b>  | <b>Analýza obtížných nahrávek</b>  | <b>Autentizace nahrávky</b>   |
|---|--|--|---|
| <ul style="list-style-type: none"><li>• kdo to mluvil?</li></ul>                      | <ul style="list-style-type: none"><li>• co nám řeč řekne o mluvčím?</li></ul>            | <ul style="list-style-type: none"><li>• co bylo řečeno?</li></ul>        | <ul style="list-style-type: none"><li>• manipuloval někdo s nahrávkou?</li></ul>      |
| <ul style="list-style-type: none"><li>• existuje nahrávka, máme podezřelého</li></ul> | <ul style="list-style-type: none"><li>• existuje nahrávka, ale žádný podezřelý</li></ul> | <ul style="list-style-type: none"><li>• nahrávka je nezřetelná</li></ul> | <ul style="list-style-type: none"><li>• chceme ověřit autentičnost nahrávky</li></ul> |

## Identifikace mluvčího



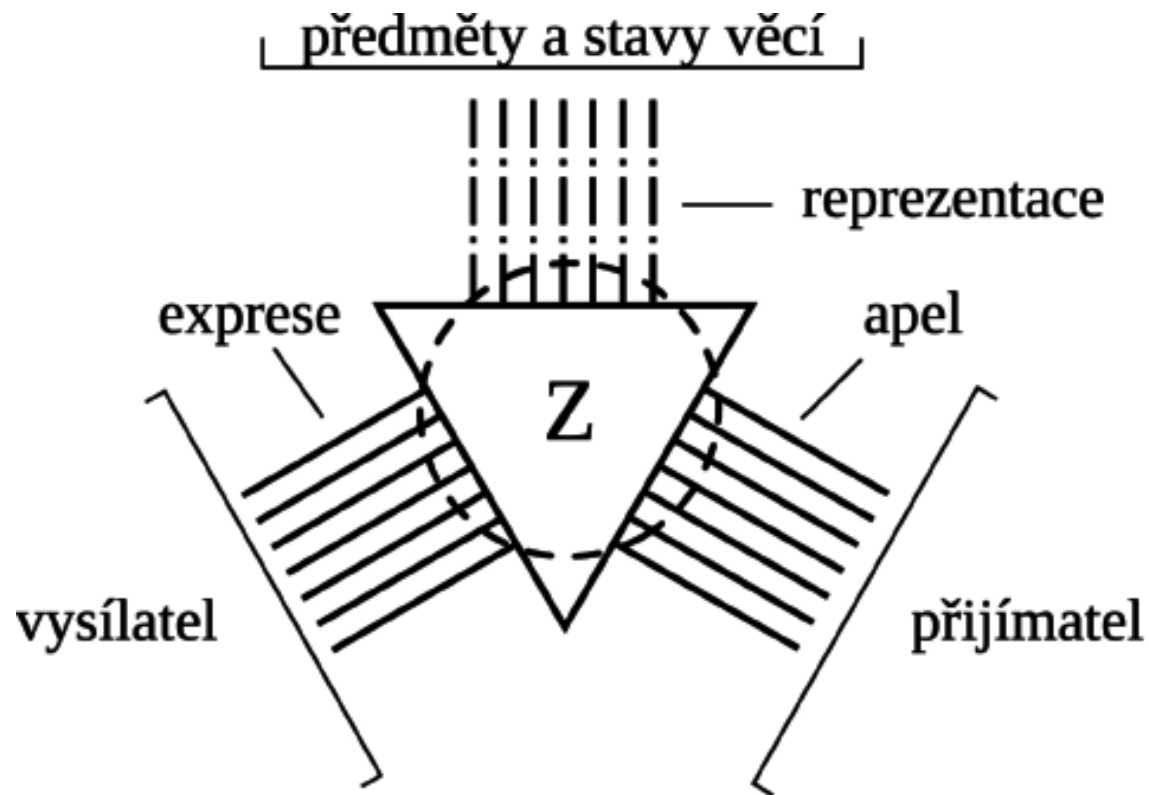
### Problémy

- Nahrávka pachatele často bývá nekvalitní
- Pachatel záměrně mění hlas
- Podezřelý je nahráván v jiné řečové situaci



## Profilování mluvčího

- Promluvy nepřenášejí jen informaci zakódovanou do slov (**reprezentativní fce**), ale také informace o nás (**expresivní/výrazová fce**)
- Promluvy zároveň nějak působí na posluchače (**apelativní fce**)
- Na našem způsobu mluvení se projevují různé sociální faktory (pohlaví<sup>[\*]</sup>, věk<sup>[\*\*]</sup>, místo narození/pobytu, vzdělání, společenské postavení, zaměstnání atd. → předmět zájmu **sociofonetiky**, odvětví **sociolingvistiky**)



[\*] biologicky dané, avšak sociokulturně ovlivněné

[\*\*] neprůkazné, mluvčí mají tendenci nadhodnocovat a podhodnocovat věk

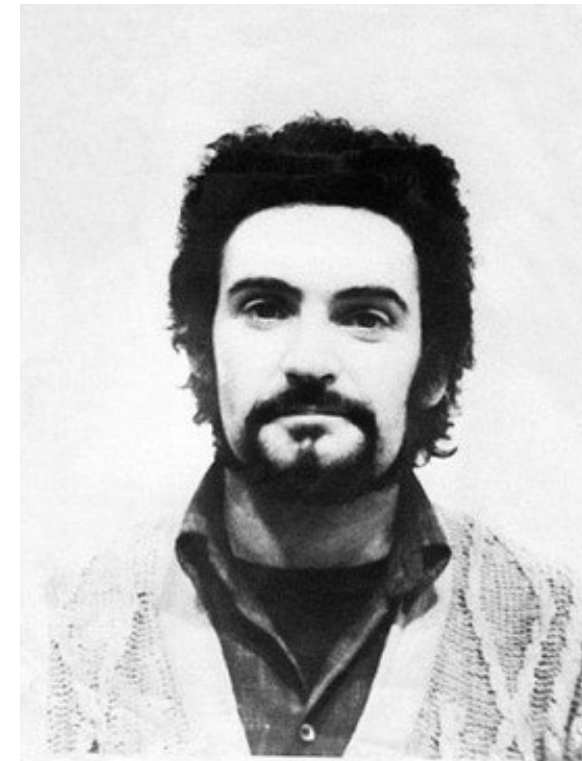
- Na mluvení se projevuje i **požití omamných látek**

- *Havárie tankeru Exxon Valdez* (největší ropná katastrofa, 1989)
  - fonetické zkoumali nahrávky kapitánových radiových zpráv
  - mohl být pod vlivem alkoholu (výslovnost [š] místo [s])



- Zkušený fonetik (jazykovědec) může z nahrávky „vyčíst“ informace o mluvčím a **vytipovat okruh podezřelých**

- *Yorkshirský rozparovač, 70. léta 20. století* (13 ubodání žen, 7 pokusů o vraždu)
  - na základě nahrávky jazykoví experti lokalizovali čtvrt ve městě Sunderland
  - autor nahrávky John Humble byl skutečně z této čtvrti, ale nebyl yorkshirským rozparovačem; tím byl Peter Sutcliffe →



## Literatura

- Skarnitzl, Radek – Šturm, Pavel – Volín, Jan. 2016. *Zvuková báze řečové komunikace: Fonetický a fonologický popis řeči*. Karolinum. [odtud většina fonetických obrázků]
- Skarnitzl, Radek (ed.), 2014. *Fonetická identifikace mluvího*. Karolinum.
- Ladefoged, Peter – Disner, Ferrari. 2012. *Vowels and Consonants*. Třetí vydání. Wiley-Blackwell.
- Munro, Murray J. 2021. *Applying Phonetics*. Wiley-Blackwell.
- Matoušek, Jindřich. 2017. SYNTÉZA ŘEČI. In: Petr Karlík, Marek Nekula, Jana Pleskalová (eds.), CzechEncy - Nový encyklopedický slovník češtiny.  
URL: <https://www.czechency.org/slovník/SYNT%C3%89ZA%20%C5%98E%C4%8CI>
- Nekula, Marek. 2017. FUNKCE JAZYKA. In: Petr Karlík, Marek Nekula, Jana Pleskalová (eds.), CzechEncy - Nový encyklopedický slovník češtiny.  
URL: <https://www.czechency.org/slovník/FUNKCE%20JAZYKA>