

„Dnešní řečové vědy jsou vysoce interdisciplinární,“ říká ředitel Fonetického ústavu FF UK Radek Skarnitzl

Ve dnech 7. až 11. srpna 2023 se v pražském Kongresovém centru konal International Congress of Phonetic Sciences (ICPhS), na jehož pořádání se podílel Fonetický ústav FF UK. O zajímavých tématech, která přinesl, mluví v rozhovoru ředitel ústavu a zároveň prezident kongresu doc. Radek Skarnitzl.



Co znamená pro českou fonetiku hostit mezinárodní setkání, na kterém se sešlo přes 1000 lidí?

Pro českou fonetiku, a zejména Fonetický ústav FF UK, je to obrovská čest. ICPhS je něco jako fonetická olympiáda – koná se jednou za čtyři roky, takže je to opravdu velká a významná

akce. Není to tak, že když se nedostanu letos, tak to vyjde za rok... Pro účastníky je to prestižní akce, a pro nás tím pádem taky. Kandidaturu jsme Mezinárodní fonetické asociaci (IPA) předkládali už v roce 2019, kdy Fonetický ústav slavil 100 let od svého založení. To, že jsme pořádání získali, beru jako uznání naší práce v posledních letech i uznání Fonetického ústavu jako celé instituce.

Jak konkrétně se na akci Fonetický ústav FF UK podílel?

Ústav byl hlavním garantem odborného programu. Spolu s kolegy z Filozofické fakulty Univerzity Palackého v Olomouci a z Psychologického ústavu AV ČR jsme byli zodpovědní za celou vědeckou stránku kongresu. Příprava odborného programu začala určením vědeckých oblastí, což může znít triviálně, ale dnešní řečové vědy jsou vysoce interdisciplinární, takže jsme jich nakonec měli 29 – od těch „klasických“ jako produkce a percepce řeči, akustické vlastnosti řeči, fonace a kvalita hlasu přes osvojování zvukové stránky mateřského i cizího jazyka, fonetiku ohrožených a málo zdokumentovaných jazyků až po klinickou fonetiku a řečové poruchy, forenzní fonetiku, neurolingvistiku či psycholingvistiku.

Dále jsme se podíleli na recenzi došlých příspěvků – je důležité zdůraznit, že příspěvky na ICPHS nejsou abstrakty jako na mnoha konferencích, ale pětistránkové svébytné studie. A dále pak zejména na finálním výběru příspěvků pro ústní a posterovou prezentaci a na sestavení celého programu. Takže Fonetický ústav FF UK u toho byl od začátku až do samého konce.

Jak se skládá program pro stovky účastníků takto interdisciplinárního kongresu?

Docela obtížně. Je těžké vyhnout se překryvům zájmu. Ale vycházel jsem hodně z vlastní zkušenosti – sám jsem tematicky hodně „široce rozkročen“, zajímá mě prozodie řeči, osvojování cizího jazyka, forenzní fonetika, sociolingvistika, nyní se

začínám věnovat i klinické fonetice. Doufal jsem, že když nebudu sám chtít vidět tři prezentace ve stejný čas, většina účastníků snad taky ne. Ale úplně se překryvům vyhnout nelze, protože po celou dobu kongresu běželo pět paralelních sekcí.

Na hlavním programu byly prezentace mezinárodních odborníků. Co zajímavého přinesly? Jaký výstup vás nejvíce zaujal?

Vrcholem pro mě byly jednoznačně tři plenární přednášky. John Esling (University of Victoria) se celý život věnuje činnosti hrtanu a různým kvalitám hlasu a jeho přednáška vlastně shrnula jeho celoživotní výzkum. Ukázky chování hlasivek v různých situacích – od zvláštních hlásek v exotických jazycích, přes operní zpěv a mongolský rap až po death metal – a vůbec složitost hrtanového chování je fascinující. Andrea Ravignani (Sapienza University of Rome) je držitelem prestižního grantu ERC a věnuje se sklonu k rytmičnosti a osvojování rytmického chování u různých živočichů, v poslední době zejména u mláďat tuleňů. A konečně Pavel Trofimovich (Concordia University) nás provedl svou vědeckou cestou v oblasti osvojování cizího jazyka a srozumitelné výslovnosti a zaměřil se na společenskou odpovědnost vědy a výzkumu obecně.

Všechny tyto přednášky byly velmi inspirativní, ale samozřejmě mě zaujala i spousta standardních příspěvků a posterů – jak studovat intonaci v přirozených a nikoli laboratorních kontextech; proč některým mluvčím rozumíme při jinak konstantních podmínkách snadněji než jiným; jak se liší „funkčně ekvivalentní“ artikulační pohyby u jednotlivých mluvčích v závislosti na jejich fyziologických vlastnostech; nebo co dělají s hrtanem beatboxeři a jak to vypadá na záznamu z magnetické rezonance.

Letošní ročník nesl podtitul „Intermingling Communities and Changing Cultures“. Jak řečové vědy reflektují změny dynamicky měnící se společnosti v souvislosti s migrací a globalizací? Jak fonetika reaguje na nové výzvy spojené s AI?

Už z výčtu témat výše vyplývá, že se řečové vědy společensky relevantním tématům věnují. V době, kdy je multilingvismus normou spíše než výjimkou, ověřujeme, jak co nejefektivněji naučit výslovnost cizího jazyka, aby řeč byla snadno srozumitelná. Fonetici se zabývají dokumentací zvukových vlastností ohrožených jazyků, podílíme se na výzkumech zaměřených například na včasnou detekci logopedických poruch u dětí nebo neurodegenerativních onemocnění z řeči starších osob.

Mnozí kolegové zkoumají sociální aspekty používání jazyka, např. genderově podmíněné rozdíly v řeči. S nástupem AI a klonování hlasů se nejen řečovní inženýři, ale i fonetici věnují detekci spoofingu, tzn. odhalování, kdy danou promluvu skutečně pronesl daný mluvčí a kdy jde o podvrh syntetizovaný počítačem. Všechna tato témata se na letošním ICPHS objevila.

Byl součástí konference také nějaký společenský a kulturní program?

Ano, účastníci měli možnost se seznámit s Prahou a povečeřet v Obecním domě. Dalším bodem programu pro zájemce byl Run for Fun kolem Vyšehradu, na který se registrovalo okolo 100 účastníků. Jen se běh konal zrovna ve středu ráno, kdy byl ten nejsilnější déšť, ale i tak běželo kolem 40 účastníků.

Praha měla tu čest hostit kongres již podruhé. Poprvé se zde fonetici setkali už v roce 1967. Jak konference tehdy probíhala?

Průběh příprav a samotné organizace popisuje závěrečná zpráva organizačního výboru, která dokládá zejména narůstající zájem účastníků a potřebu několikrát navýšit rozpočet akce. Zajímavé je, že ačkoliv přípravy zabraly něco přes dva roky, korespondence Bohuslava Hály, tehdejšího ředitele Fonetického ústavu FF UK, dokládá, že za pořadatelství v Praze Hála lobboval už od roku 1956, tedy přes 10 let. O tom píše i ve svém historickém příspěvku kolega z Fonetického ústavu FF UK

dr. Pavel Šturm.

K pražskému kongresu máme také vzpomínky účastníků, které se mu podařilo dohledat – kromě profesorky Palkové jde o několik zahraničních hostů, kteří se shodují, že setkání tehdy mělo ještě „rodinnou“ atmosféru a skutečné renomé ve smyslu odbornosti a důležitosti pro obor. Ohlasy nalezneme také v denním tisku, vyšel asi tucet zpráv s úsměvnými titulky jako „800 Higginsů zasedá v Praze“.

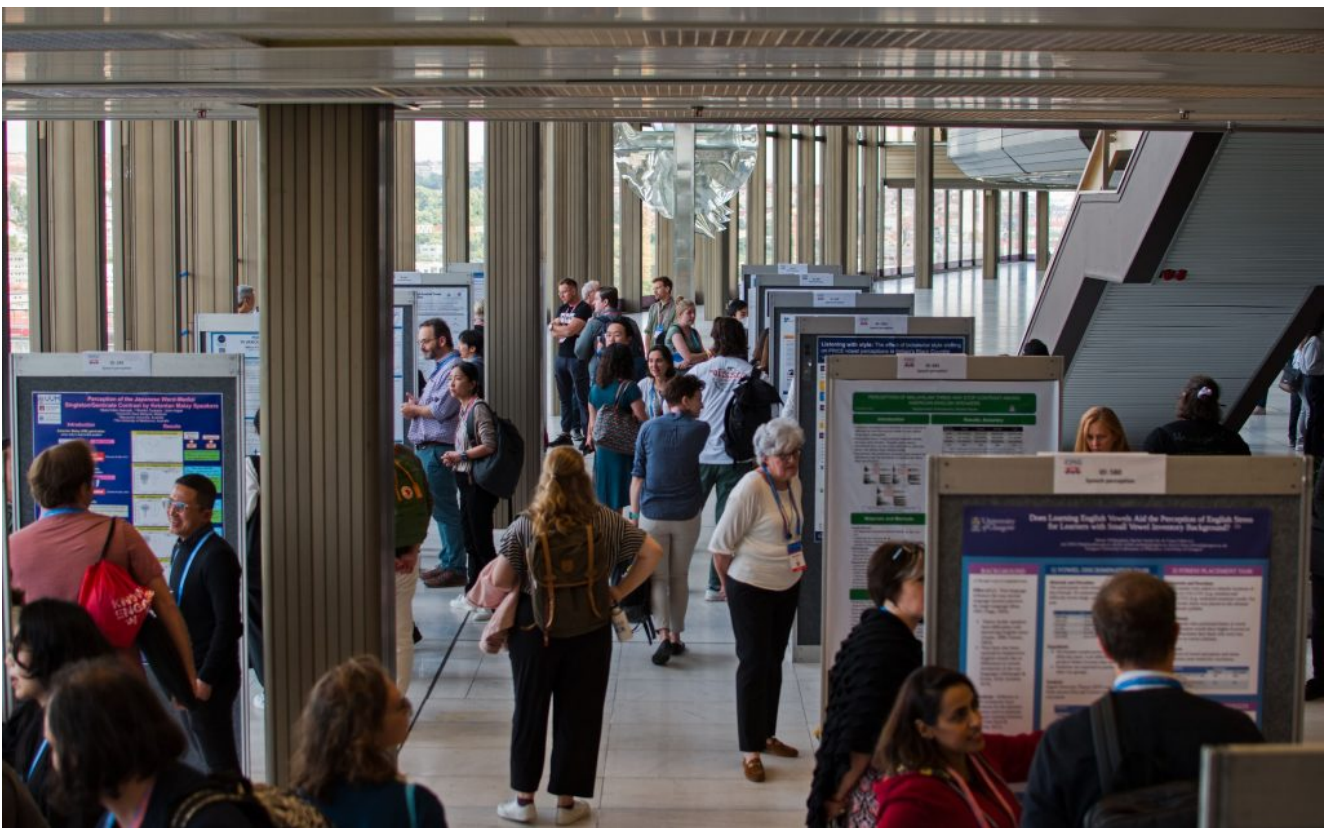
Fonetika se od té doby enormně posunula. V čem nejvíce?

Hlavně se samozřejmě znásobily možnosti, jak vůbec řeč můžeme zkoumat. V roce 1967 se spektrogram (stručně řečeno průběh kvality řečového zvuku v čase) získával velmi náročným a zdoluhavým způsobem, dnes ho v počítači vidíme okamžitě. Akustická analýza je tak velmi rychlá a možnosti se stále rozvíjejí – ostatně jedním z přednášejících na kongresu byl Paul Boersma, hlavní autor programu Praat, který k akustické analýze řeči používají fonetici po celém světě. Kromě toho existují různé možnosti, jak zkoumat přímo artikulační pohyby – vrchol představuje záznam magnetické rezonance celého mluvního ústrojí v reálném čase, o které jsem se už zmínil. To je obrovský posun od fotografování jednoho doteku jazyka na umělé patro v 60. letech. A je zřejmé, že všechny tyto možnosti hranice našeho poznání o produkci a percepce řeči posouvají, i když ještě zdaleka nevíme všechno.

Kde se uskuteční ICPHS 2027 a jak se na příští konání těšíte vy sám?

Další pořadatel se vybírá na stávajícím kongresu. Tentokrát přišla jen jedna nabídka, takže výběr nebyl složitý a 21. kongres se bude konat v kanadské Victorii. Rok 2027 je daleko, takže uvidíme i podle dostupnosti financování, ale samozřejmě bych se kongresu zúčastnil rád, protože je to opravdu jedinečná akce, nejen z hlediska vědeckého, ale i díky setkávání s kolegy a přáteli.

Štěpánka Zelenková Semecká



PRAK: AN AUTOMATIC PHONETIC ALIGNMENT TOOL FOR CZECH

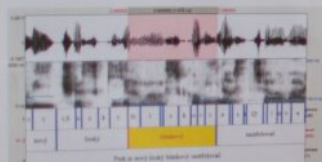
Václav Hanzl¹, Adléta Hanzlová²

¹Freelance researcher, ²Institute of Phonetics, Charles University in Prague
vhanz@pau1.com, adleta.hanzlova@pau1.com

Abstract

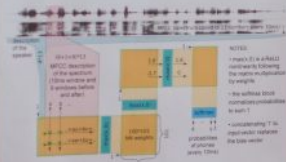
Labeling speech down to the identity and time boundaries of phones is a labor-intensive part of phonetic research. To simplify this work, we created a **free open-source tool** generating **phone sequences** from Czech text and **time-aligning** them with audio. Low architecture complexity makes the design approachable for students of phonetics. Acoustic model (AM) NN with 36k weights was trained using **PyTorch** on small CommonVoice data. Alignment and **variant selection** decoder is implemented in **Python** with matrix library. A Czech pronunciation generator is composed of simple rule-based blocks capturing the logic of the language where possible, allowing modification of transcription approach details. Compared to tools used and data preparation efficiency improved, the tool is usable on **Mac, Linux and Windows** in **Praat GUI** or command line, achieves mostly correct pronunciation variant choice including **glottal stop** detection, algorithmically captures most of Czech assimilation logic and is both discrete and practical.

Praak used in Praat GUI



Praak is fully integrated in Praat - select a sound and brought with input text and run our script. Behind the scenes, the Python-based aligner is run and the result is shown back in Praat. The script also does many sanity checks and allows several modes, e.g. alignment of multiple sound+phone couples or using one input text/prg for multiple sounds.

DNN Acoustic Model of Phones



For each blue frame, DNN AM computes probabilities of phones. Viterbi decoder then finds the best alignment for a message graph from the pronunciation generator.

Download Praak

<https://github.com/vaclavhanzl/prak>



Software, documentation, README, articles, discussions forum, slides, references, this poster as pdf. The README explains installation for all platforms.

Praak Design Goals

- an open-source tool free for any use - MIT license for code, trained on free audio data
- functioning on Mac, Linux and Windows
- easy to install - low dependencies, only reliable dependencies which are (hopefully) here to stay
- simple architecture, preferably building on techniques from phonetics students' curricula
- usable from both GUI and command line
- using explicable and modifiable logic (rather than a trained blackbox) where possible
- automatic pronunciation variant selection

Pronunciation Generator

String replacement rules cover foreign words. The regular part of Czech pronunciation is handled by a series of Finite State Transducers processing the phone string backward, taking care of assimilation or optional glottal stop insertion, for example:



Command Line prongen Tool

For a given Czech text, the pronunciation generator creates a "message graph" suggesting variants of some phones (here in blue, these are resolved by the AM).

```
prongen --text "prongen" --output "prongen" --am "am" --viterbi "viterbi" --glottal "glottal" --assimilation "assimilation" --rules "rules" --help
```

DNN Acoustic Model Training

Alignment and model are gradually built in tandem by repeated training of the AM on the previously aligned data and realigning the same data in the new version of

Both replacement rules and FSTs allow generation of multiple variants. These are resolved by the Acoustic Model match.

Precision

Percentage of phone mismatch and

Alignment Emerging in Training

Each line in phone alignment is one train-align cycle









Foto: Garant International, s. r. o.