

# Co na srdci, to na jazyku?

Jan Pešán

Speech@FIT - Fakulta informačních technologií, Vysoké učení technické v Brně

Email: ipesan@fit.vutbr.cz

Vojtěch Juřík

Psychologický ústav, Filozofická fakulta Masarykovy univerzity, Brno

Email: jurik.vojtech@mail.muni.cz

## Abstrakt

Identifikace kognitivního nebo fyzického přetížení je zásadní v řadě oblastí, kde lidské rozhodování představuje klíčový faktor týkající se ochrany zdraví a bezpečnosti osob či majetku. Specificky se jedná o oblasti, ve kterých se pohybují lidé jako piloti stíhacích letounů, řidiči autobusů, chirurgové či operátoři jaderných elektráren. Okamžitá stresová reakce na fyzické nebo kognitivní podněty může zásadně narušit fungování člověka a ztížit tak proces rozhodování v kritických situacích. Význam paralingvistického automatického zpracování řeči zde tudíž nelze opomíjet. Intenzita, úroveň vyčerpání, intonace a rytmus řeči jsou příklady paralingvistických vlastností mluveného projevu, které mohou nejen změnit význam komunikovaného obsahu, ale mohou také posloužit jako příznaky vhodné pro detekci rizikových situací. V rámci výzkumu je třeba věnovat tímto paralingvistickým vlastnostem pozornost a snažit se vyvinout nástroje pro jejich efektivní rozpoznávání. Nicméně, problém pro celou oblast stále představuje nedostatek vysoce kvalitních referenčních dat pro trénování detekčních systémů. Jako příspěvek k řešení této otázky jsme vyvinuli nástroj BESST pro generování potřebných řečových dat ve stresovém kontextu a s jeho pomocí jsme shromáždili reálná data mapující stresové projevy v lidské řeči. Tento příspěvek diskutuje možnosti a omezení vyplývající z navrženého nástroje, k čemuž využívá analýzy originálních naměřených dat.

## 1 Stres a řeč

Účinky lidského stresu na produkci řeči mohou být významné. Stres může způsobit změny v hlasu, jako je chrapot, třes a změny výšky a hlasitosti (Laukka a spol., 2008), (Zsiga, 2013). Tyto změny jsou způsobeny fyziologickou reakcí těla na stres, která může zahrnovat zvýšené napětí ve svalectech používaných pro řeč. Když je tělo ve stresu, aktivuje se sympatický nervový systém (SNS), což může způsobit uvolnění adrenalinu a kortizolu (Dickerson a Kemeny, 2004). Adrenalin může zvýšit svalové napětí, zatímco kortizol může ovlivnit

hlas tím, že způsobí vysychání hlasivek. To může vést ke změnám výšky a hlasitosti, stejně jako ke zvýšení hlasového úsilí, což může dále zhoršit napětí ve svalectech používaných pro tvorbu řeči. Kromě změn v hlase může stres ovlivnit také kognitivní procesy, které se podílejí na tvorbě řeči. Pro tělo ve stresu, může být obtížné soustředit se, což může ovlivnit schopnost plánovat a organizovat řeč. To může mít za následek neorganizovanou nebo obtížně sledovatelnou řeč (Laures-Gore a spol., 2019). Stres může navíc způsobit změny v rychlosti a rytmu řeči. Lidé ve stresu často mluví rychleji nebo koktají, což může ostatním ztěžovat porozumění. To může vést k chybné komunikaci a dalšímu zvýšení úrovně stresu, což lze rovněž považovat za další faktory, na základě kterých je možné identifikovat stres u člověka.

### 1.1 BESST - Brno Extended Stress and Speech Test

Brno Extended Stress and Speech Test (BESST) je experimentální metodologie určená pro sběr dat, která jsou vhodná pro trénování strojového učení (ML) pro detekci stresu v řeči. Metoda je navržena tak, aby maximalizovala sběr řečových výstupů od účastníků umístěných v náročných stresových situacích. K tomuto účelu bylo navrženo experimentální prostředí, které tvoří záznam z několika kamer a mikrofonů v situacích, kdy účastníci plní řadu úkolů, které mají stresový charakter. V prvním úkolu je fyziologický stres vyvoláván ponořením nedominantní ruky účastníka do ledové vody, zatímco je zaznamenáván jeho volný projev řeči. Druhý stresový kontext je založen na duálním úkolu, který využívá tzv. Reading Span Task Daneman a Carpenter (1980), který zvyšuje kognitivní zátěž, zatímco účastník řeší a čte textový rébus.

Kromě audio a video záznamů jsou také pořizovány elektrokardiogram (EKG) pomocí zařízení Faros 180<sup>1</sup> a měření hladiny galvanického odporu kůže (GSR) pomocí náramku Empatica E4<sup>2</sup>. Data z obou senzorů nám pomáhají hodnotit objektivní úroveň fyziolo-

<sup>1</sup><https://www.bittium.com/medical/bittium-faros>

<sup>2</sup><https://www.empatica.com/research/e4/>

gického stresu. Subjektivní stavy stresu jsou účastníky hlášeny prostřednictvím Perceived Stress Scale 14 (PSS14; Cohen a spol., 1983), State Trait Anxiety Inventory Y1 (STAI-Y1; Hedberg, 1972) a NASA Task Load Experience (NASA TLX; Hart a Staveland, 1988), a jsou také hodnoceny na základě vlastních odhadů zátěže po každém úkolu. Celá metodologie je navržena tak, aby byla funkční a škálovatelná pro sběr klíčových dat nezbytných pro detekci stresu pomocí strojového učení a mohla tak být přenesena do jiných výzkumných prostředí nebo kulturních kontextů, čímž nabízí také možnost sběru dat na různých populacích.

## 1.2 Dosavadní závěry

V rámci studie jsme se zaměřili na vztah mezi srdeční činností a subjektivním/objektivním stresem. Zjistili jsme, že srdce reaguje relativně rychle na stresový podnět, ale jakmile tělo vyhodnotí míru ohrožení, adaptuje se velice rychle zpět (mezi 20-30 úderů srdce), což komplikuje využití frekvence srdečního rytmu jako objektivní indikátor stresu. Tato komplikace má vliv na analýzu řeči - je potřeba zachytit akutní stresovou reakci zároveň s řečí (ale 20-30 úderů je pouze 15-30 sekund, což představuje pro potřebnou analýzu malý překryv).

Další výzvou se ukazuje správné zarovnání stresové reakce a řečových projevů v čase. I malý posun v časování může být problematický a vést k nesprávné interpretaci dat. Kromě toho je třeba zvážit, jaká granularita dat je nejvhodnější pro jejich anotaci - zda se jedná o celé experimenty, části experimentů nebo dokonce mikrosegmenty.

### 1.2.1 Synchronizace datových zdrojů

V experimentu BESST jsme použili několik nezávislých datových zdrojů; audiovizuální záznamy, EKG, GSR. Abychom mohli tato data správně analyzovat a interpretovat, je důležité zajistit, aby byla synchronizována s přesným časováním. Existuje několik přístupů, jak uvádíme v textu dále.

### 1.2.2 Explicitní synchronizace

Jednou z možností, jak zajistit synchronizaci, je použití tzv. explicitní synchronizace. To znamená, že v určitém okamžiku vytvoříme viditelnou a slyšitelnou zvukovou událost (např. tlesknutí), na základě kterého jsme schopni data synchronizovat, protože slouží jako referenční bod pro synchronizaci dat. Nicméně, u některých datových zdrojů, jako je EKG, je obtížné použít tuto metodu.

### 1.2.3 Implicitní synchronizace pomocí časových značek

Další možností je použití implicitní synchronizace pomocí časových značek. Každý datový zdroj může být označen časovou známkou, která určuje přesný čas, kdy byl zdroj zaznamenán. Tyto časové značky se následně používají k synchronizaci dat. Nicméně, časové značky mohou driftovat s ohledem na použitá zařízení, takže je důležité dbát na to, aby byly zařízení správně kalibrovány a aby nedošlo k rozdílnému driftu u různých zařízení.

### 1.2.4 Lab Streaming Layer

Další možností je použití nástroje Lab Streaming Layer (LSL)<sup>3</sup>, což je open-source nástroj pro synchronizaci datových zdrojů. LSL umožňuje snadnou integraci různých datových zdrojů a poskytuje přesné časování a synchronizaci. Nicméně, použití LSL vyžaduje určitou konfiguraci a způsob použití, takže jeho nasazení není v některých kontextech vhodné.

## 2 Závěr

Výzkum stresu a jeho detekce pomocí řečových projevů a strojového učení přináší řadu výzev, a to jak na úrovni metodologické, tak na úrovni analýzy dat. Zároveň nabízí značný potenciál pro celou diskutovanou oblast, kdy se hledání cest k dosažení validních dat vhodných pro strojové učení stává středobodem celé problematiky. V tomto ohledu dále rozvíjíme metodologii BESST a shromažďujeme data o lidském řečovém projevu naměřená v rámci stresového kontextu. Aktuálně plánujeme uveřejnění prvního datasetu pro účely strojového učení a hledáme adekvátní způsoby jejich analýzy. Součástí plánované publikace dat je podrobný popis metodologie sběru dat, jejich podrobná charakteristika, návrh možného zpracování, popis limitů celého výzkumu a doporučení ve formě nabýtvých zkušeností.

Souhrnem, navrhovaná metodologie BESST poskytuje funkční a škálovatelnou základnu pro sběr klíčových dat nezbytných pro detekci stresu pomocí hlubokého učení, a to nejen v českém prostředí, ale i v jiných kulturních a jazykových kontextech.

## Reference

Cohen, S., Kamarck, T. a Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24(4):385.

Daneman, M. a Carpenter, P. A. (1980). Individual di-

<sup>3</sup><https://labstreaminglayer.org/>

ferences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4):450–466.

Dickerson, S. S. a Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, 130(3):355–391.

Hart, S. G. a Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. V *Advances in Psychology*, str. 139–183. Elsevier.

Hedberg, A. G. (1972). Review of state-trait anxiety inventory. *Professional Psychology*, 3(4):389–390.

Laukka, P., Linnman, C., Åhs, F., Pissiota, A., Frans, O., Faria, V., Michelgard Palmquist, A., Appel, L., Fredrikson, M. a Furmark, T. (2008). In a nervous voice: Acoustic analysis and perception of anxiety in social phobics' speech. *Journal of Nonverbal Behavior*, 32:195–214.

Laures-Gore, J., Cahana-Amitay, D. a Buchanan, T. W. (2019). Diurnal cortisol dynamics, perceived stress, and language production in aphasia. *Journal of Speech, Language, and Hearing Research*, 62(5):1416–1426.

Zsiga, E. (2013). *The sounds of language : an introduction to phonetics and phonology*. Wiley-Blackwell, Chichester.