



FILOZOFICKÁ FAKULTA
UNIVERZITY KARLOVY
V PRAZE



FONETICKÝ ÚSTAV

DIPLOMOVÁ PRÁCE

David Lukeš

Percepční citlivost ve frekvenční a temporální doméně u hudebních a řečových stimulů

Perceptual sensitivity to music and speech stimuli in the
frequency and temporal domains

Na tomto místě bych rád poděkoval Janu Volínovi za trpělivost, pevné vedení, schopnost usměrnit mou těkavou pozornost, a v neposlední řadě i za návrh zajímavého tématu a nesmírně přínosné konzultace k němu. Velmi si vážím také času, který výzkumu věnovali respondenti pilotáže i ostré fáze percepčního testu. A konečně patří dík i Filozofické fakultě Univerzity Karlovy, která mi prostřednictvím účelového stipendia přispěla na účast na *International Summerschool on Agent-based Models of Creativity v Cortoně* (září 2013): zejména odd. 2.1 těží z poznatků, které jsem tam o vztahu řeči a hudby získal.

Prohlašuji, že jsem diplomovou práci vypracoval samostatně, že jsem řádně citoval všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokéhoškolského studia či k získání jiného nebo stejného titulu.

V dne

podpis:

Abstrakt

Abstract

Obsah

Abstrakt	3
Abstract	3
Úvod	7
I Teoretické pozadí	9
1 Hudba a řeč: od metafory k neurofyzilogickému propojení	10
1.1 Zvuková charakteristika hudby	14
1.2 Zvuková charakteristika řeči	15
2 Spřízněné domény?	18
2.1 Evoluční pohled	18
2.2 Výzkum temporální a frekvenční struktury řeči a hudby	26
2.2.1 Hudba a řeč v mozku	27
2.2.2 Kvantifikace strukturních podobností mezi řečí a hudbou	33
2.2.3 Právě postřehnutelné temporální a frekvenční rozdíly (JND)	41
2.2.4 Řeč a hudba v percepci	42
II Experiment	50
3 Výzkumné paradigma	51
4 Metoda	53
4.1 Řečové stimuly	53
4.2 Hudební stimuly	56
4.3 Hudební stimuly pro smíšené položky	58
4.4 Pilotáž a sestavení testu	59
4.5 Ostré testování	61
5 Výsledky	64
5.1 Charakteristika respondentů	64
5.2 Konzistence a správnost odpovědí	64
5.3 Korelační měření	71
5.4 Lineární modely: hledání faktorů úspěšnosti	76

6	Diskuse	82
6.1	Vliv kategorických proměnných <i>hudebnost</i> a <i>VŠ</i>	82
6.2	Provázání řečové a hudební percepce	83
6.3	Metoda a další výzkum	85
	Závěr	91
	Bibliografie	93

Information is not knowledge.

Knowledge is not wisdom. Wisdom is not truth. Truth is not beauty.

Beauty is not love. Love is not music.

Music is the best...

Frank Zappa, „Packard Goose“

Úvod

Předkládaná diplomová práce se zabývá percepční citlivostí na drobné variace ve frekvenční a temporální výstavbě řečových a hudebních stimulů a případné korelace mezi výkonem posluchačů v obou doménách. V první, teoretické části nabízí přehled současného stavu poznání v oblasti výzkumu paralel mezi jazykem/řečí a hudbou. Jak dokazuje relativně nedávná Patelova monografie (2008), jde o dynamickou oblast bádání, která se snaží o interdisciplinární propojení výzkumných výsledků z různých vědeckých sfér, a slibuje tak komplexní poznatky na pomezí fonetiky, lingvistiky, hudební vědy, psychologie a neurověd, ba i evoluční antropologie a filozofie.

Text postupuje od obecného úvodu o vztahu mezi řečí a hudbou, spojeného s krátkým pohledem do historie, přes porovnání charakteristických postupů při výstavbě zvukové struktury v obou doménách, až k detailnímu zhodnocení důkazů, které hovoří buď pro nebo proti spříznění obou domén. Toto spříznění na sebe může vzít mnoho podob, od evolučního propojení přes sdílené neurální struktury až po kulturně podmíněné paralely mezi frekvenčními a temporálními vzorci v řazení řečových a hudebních jednotek. Zvláštní pozornost je pak věnována percepčnímu přístupu k problematice, v němž spočívá metodologie i této práce. Ten sice na rozdíl od neurověd neumožňuje činit silné závěry o biologicko-fyziologickém aspektu věci, ale na druhou stranu tím, že vychází z chování subjektů, zahrnuje vědomou interakci s okolním světem. Je tudíž větší apriorní pravděpodobnost, že případné objevené efekty mají reálný dopad v životě jedince, a nejedná se jen o interní detaily „zapojení“ jeho mozku, které se na povrchu setrou.

Těžištěm práce je tedy empirická studie, v níž měli respondenti posuzovat jemné rozdíly ve třech typech stimulů: řečových, hudebních a smíšených (řečově-hudebních). Průzkumem takto získaných dat se budeme snažit přijít na to, zda jsou výkony subjektů v jednotlivých částech testu korelované, a zda je tedy možné se domnívat, že jsou hudební a řečová percepční citlivost do jisté míry provázané, či zda úspěšnost souvisí spíše s jinými faktory, jako je celková hudební zkušenost či lingvisticko-fonetické vzdělání.

V kap. 1 poskytneme úvodní pohled na různé paralely mezi řečí a hudbou, od těch čistě metaforických přes jejich společenskou funkci až po možnou neurofyziologickou integraci. Také podáváme stručnou zvukovou charakteristiku obou domén, přičemž si všímáme, že zatímco hudba bývá převážně organizována okolo kontrastů výšky (frekvence), řeč naopak okolo kontrastů tónu (spektra).

Kap. 2 pak představuje hlavní přehled dosavadních poznatků o spřízněnosti obou domén. Obsahuje příklady ze tří důležitých okruhů experimentálních paradigmat, jimiž se hypotézy o tomto vztahu zkoumají. V evoluční biologii (odd. 2.1) jde o usouvztažnění mechanismů, které vedly k vzniku obou systémů; na pomoc si bere srovnání člověka s dalšími druhy zvířat schopnými komplexních vokalizací. Neurovědy (odd. 2.2.1) zkoumají

odraz zpracovávání řeči a hudby přímo v mozku pomocí zobrazovacích metod, jako je např. fMRI, a jsou tedy schopny vyhodnotit aktivace neuronů spojené s řečí a hudbou. Důležitou roli percepčních experimentů (odd. 2.2.4) jsme popsali již výše.

Kapitola se dále zabývá možnostmi kvantifikace strukturních podobností ve výstavbě řeči a hudby, která přináší zajímavé poznatky o kulturní podmíněnosti hudebních tradic (odd. 2.2.2). Odd. 2.2.3 je pak věnován právě postřehnutelným rozdílům ve frekvenčních a temporálních charakteristikách zvuku, tj. hranicím lidské percepční citlivosti, a rozebírá relevanci těchto hodnot pro náš experiment.

Experimentální část práce je uvozena krátkou charakteristikou použitého výzkumného paradigmatu a připomenutím výše popsaných jednoduchých korelačních hypotéz, s nimiž do experimentu vstupujeme (kap. 3). Následuje detailní popis metody (kap. 4): od tvorby stimulů (odd. 4.1–4.3), přes rozbor výsledků pilotáže a sestavení testu (odd. 4.4) až po samotné ostré testování.

V kap. 5 shrnujeme výsledky percepčního testu. Podáváme charakteristiku vzorku respondentů (odd. 5.1) a ověřujeme, do jaké míry byli v plnění úloh konzistentní a úspěšní (odd. 5.2). Korelace mezi jednotlivými částmi testy jsou převážně statisticky významné, síla jejich efektu se pohybuje od nízké po středně silnou (odd. 5.3); nejsilnější je propojení mezi skóre v hudební a smíšené části. Pomocí lineárních modelů pak zkoumáme v kontextu výkonů v jednotlivých částech testu i vliv dalších faktorů: předchozí hudební zkušenosti a jazykového vzdělání (odd. 5.4). Zatímco první z nich v některých ohledech roli hraje, druhý s výkonem v testu nijak nesouvisí.

V závěrečné kap. 6 a snažíme se o komplexnější interpretaci pozorovaných jevů ve světle poznatků shromážděných v teoretické části, zejm. z odd. 2.2.4. Jedním z důležitých témat je úvaha nad možnými příčinami toho, že nebyly pozorované korelace silnější, a podáváme návrhy na budoucí výzkum, který by mohl spřízněnost percepční citlivosti na řeč a hudbu popsat robustněji (odd. 6.3).

Everything on this planet has
something to do with music. Music
functions in the realm of sculptured
air.

Frank Zappa, duben 1979

Část I

Teoretické pozadí

1 Hudba a řeč: od metafory k neurofyziologickému propojení

Na vztah mezi hudbou a jazykem lze nahlížet z celé škály perspektiv, od čistě metaforické po ryze materiální, např. akustickou či biologickou. Na **metaforické** spříznění hudby a jazyka je snadné poukázat. Oba jsou to při pohledu zvenčí znakové systémy, podléhají jistým konvencím a jsou lidskými univerzáliemi (Patel 2008, s. 3). Složitá struktura jazyka jakožto komunikačního prostředku je pro lidský druh zcela specifická; stejně tak hudba v přísném slova smyslu je podle Raye Jackendoffa vlastní všem lidským kulturám, a přitom právě jenom jim: „žádný živočich kromě člověka nevykazuje schopnost jazyka či hudby v lidském slova smyslu, a co víc, u ostatních primátů nelze nalézt ani přímočaré evoluční předchůdce obou těchto schopností“ (2009, s. 196; viz též Patel 2008, s. 3 aj.).¹ Dokonce i pokud pojmem hudební dovednost trochu volněji jako schopnost vokální imitace a motivické práce (přeskupení, obměny) s naučenými sekvencemi, je sdílena poměrně úzkým okruhem živočichů. Kromě lidí jde např. o zpěvné ptáky a některé kytovce, ale ostatní primáti jsou znovu vynecháni (Okanoya et al. 2007, s. 363–6). K tomu ještě dodejme, že jak jazyk, tak hudba využívají pro přenos informací stejné médium (zvuk) a v neposlední řadě se alespoň částečně překrývají jejich běžná pole působnosti, a to ve schopnosti ovlivňovat afektivní stavy posluchačů či účastníků konverzace.

Z hlediska metafor lze mluvit o „hudebním jazyku“ daného skladatele (myšleno soubor jeho charakteristických výrazových postupů) či naopak o „hudbě něčí řeči.“ Některá tato spojení jsou dokonce lexikalizovaná, srov. věta v řeči a věta v orchestrální skladbě. Metafora ovšem klouzájí po povrchu, stačí jim vnější, dílčí podobnosti, jež jsou sice zajímavé z hlediska vztahů v konceptuálním sémantickém prostoru, který si vytváříme, ale nemusí nic vypovídat o genezi či systematické spřízněnosti jevů, jež popisují. Ony podobnosti jsou navíc silně abstraktní, z čehož plyne, že mají velkou šíři záběru, a shody, které zachycují, nemusí být natolik jedinečné, aby byly významné. Kromě „hudebního jazyka“ tak můžeme hovořit např. o „architektonickém jazyku,“ aniž bychom tím mínili postulovat nějakou hlubokou souvislost mezi řečí a stavebnictvím.

Již konkrétnější a lépe ověřitelná jsou tvrzení o jisté **funkční ekvivalenci** mezi jazykem a hudbou, ač k metaforičnosti pochopitelně příležitostně svádějí či sklouzávají. Často vycházejí právě z výše zmíněné manipulace s afektivními stavy, která je primárním účinkem hudby, ale i v jazykové komunikaci může hrát důležitější roli než samotný obsah sdělení. V tomto viděli významnou úlohu hudby mnozí skladatelé i filozofové 19. stol., mezi nimi např. Richard Wagner a Friedrich Nietzsche. V jejich pojetí měla hudba (silněji než jakýkoli jiný výrazový prostředek včetně jazyka) nejen vyjádřit, ale přímo v poslu-

¹“[N]o nonhumans have either language or music in the human sense, and in particular there are no obvious evolutionary precursors for either in non-human primates.”

chačích probudit silné emocionální pohnutí, které se otiskne do společnosti a změní ji. Nietzsche dokonce ve *Zrození tragédie* hudbě přisoudil větší expresivní moc než jazyku: „Proti [hudbě] každý jev jest pouhým podobenstvím: proto *jazyk*, jakožto nástroj a náznak jevů, nikde a nikdy nemůže navenek vynéstí nejhlubšího nitra hudby“ (Nietzsche 2008, s. 26). Nejde ani tak o samotné nadřazení hudby jazyku: jeho záměrem je čistě rétorická valorizace hudby a samotný obsah tohoto tvrzení – že hudbu nelze vyjádřit jazykem – je banalitou, stejně jako tvrzení opačné, že jazyk nelze vyjádřit hudbou.² Důležitý je spíš postřeh, že hudba a jazyk zápolí na stejném kolbišti.

Pojetí hudby, v němž má pár jedinců zaujmout masy a pohnout jimi (což se dá snadno zneužít za účelem manipulace, viz např. jak si hitlerovské Německo přivlastnilo právě Wagnera), je typické pro současnou západní kulturu. Hudba je v ní do velké míry institucionalizovaná, spojená s autoritou, a rozděluje společnost na interprety a posluchače. Z antropologického hlediska je ovšem mnohem původnější pojetí hudby jako participativního společenského rituálu. Členové pygmejského kmene Mbendjele žijícího v pralesích Konga-Brazzaville pomocí mnohahodinových hudebních obřadů udržují pospolitost celé komunity a zároveň jejím mladším členům vštěpují její organizaci a způsob fungování. Jak vysvětluje antropolog Jerome Lewis, kmen Mbendjele je silně rovnostářský, neexistuje v něm hierarchické uspořádání, které by se staralo o koordinaci aktivit společenství za účelem obživy. Pak ovšem hrozí nebezpečí, že pokud nebudou jednotliví členové své individuální činnosti spontánně diferencovat a zároveň směřovat ku prospěchu celku, přežití bude ohroženo: „kdyby všichni šli shánět med, nemusí nakonec být nic k jídlu“ (Lewis 2012, s. 100). Právě společné hudební produkce toto tláhnutí ke spontánní spolupráci v Pygmeích posilují: jejich složitá polyfonní hudba vyžaduje svědomitou účast všech a nutí je, aby si zachovali žádoucí stupeň individuality (jinak by se polyfonie ztratila), ale přitom na svůj part nestráhali příliš mnoho pozornosti (jinak by byla ohrožena křehká polyfonní rovnováha). Pro děti se tento postoj k životu, který jim je zmíněným způsobem nenásilně vštěpován již od malička, stává druhou přirozeností a přijímají jej za svůj.

Wagnerův sen – že hudba bude mytologicko-narativním tmelem, který utuží společenství a jeho identitu (viz třeba Garratt 2010, s. 168) – je pro tyto Pygmeje každodenní realitou. Hudba se tak stává prostředkem politicko-sociálního uvědomění, což je role, kterou v západní společnosti plní prostřednictvím např. vzdělávacích institucí a médií hlavně jazyk. Až do této překvapivé míry se tedy v praxi mohou oba systémy funkčně zastupovat. Problém takto funkčně definovaného vztahu mezi hudbou a jazykem pak samozřejmě tkví v tom, že ačkoli již není tolik elastický jako metafora, stále nabízí příliš mnoho interpre-

²Ale mluvčí některých jazykových komunit (zejména u jazyků bohatě využívajících tóny jako lexikálně distinktivní prostředek) vyvinuli dorozumívací kódy, které pro vyjádření jazykových jednotek využívají hudebních elementů, např. hvízdání (v mexickém jazyce Chinantec) či bubnování (západní Afrika). Detaily viz Patel (2008, s. 48–49).

tační volnosti a umožňuje přání, aby se stalo otcem myšlenky.

O další krok blíž k vědecké ověřitelnosti (a tedy i falzifikovatelnosti) jsou teze o **evolučním propojení** hudby a jazyka. Jejich nevýhodou pochopitelně je, že k přímým pozorováním evolučních stadií obou jevů nemáme přístup, závěry musíme vyvozovat z modelů, analogií nebo „fosílií“, a těch jazyk ani hudba – jakožto původně nemateriální kulturní artefakty – z tohoto nejranějšího období příliš nezanechaly. Již do dob Charlese Darwina se datují vědecké hypotézy o fylogenetické spřízněnosti řeči a hudby; Darwin sám podal spekulativní popis tohoto domnělého procesu:

[...] dávný člověk, či spíše nějaký ještě dávnější předek člověka poprvé užil svého hlasu nejspíše tehdy, když se pokoušel o skutečné hudební kadence, tedy o zpěv, jak to v současné době můžeme pozorovat u některých gibbonů. Podle četných analogií pak můžeme soudit, že se tato schopnost rozvíjela zejména v období námluv, kdy zpěv vyjadřoval rozličné pocity, jako například lásku, žárlivost, radost z vítězství či výzvu k soupeři. (Darwin 2006, s. 106)³

Svébytnou formou výzkumu v této oblasti je vytváření počítačových simulací (tzv. agentních modelů), v nichž virtuální entity za přesně daných podmínek interagují v předem definovaném prostředí a sleduje se, které iniciální podmínky vedou ke zvolení kterých adaptivních strategií⁴ a rozvinutí souvisejících dovedností. Řeč i hudba se v takových simulacích považují za projevy komplexně adaptivního chování (*complex adaptive behavior*, nebo též *emergent behavior*), tedy chování, při němž vícero jedinců koordinuje své aktivity, aniž by je zvenčí řídil nějaký vůdce či plánovač s představou o celkovém směřování jejich snažení. Podobně jako se rybí a ptačí hejna explicitně nedomlouvají na směru pohybu, ale přesto jej mění až překvapivě plynule a jednoduše, tak lidé „za běhu“ vytvořili složitý komunikační kód (jazyk), aniž by se na něm museli předem domlouvat, a při hudební improvizaci dokážou na místě budovat společnou kompozici. Komputační modely se pak snaží prozkoumat podmínky (z hlediska kognitivní vybavenosti apod.), za jakých je možné těchto jevů docílit a sladit chování jednotlivých agentů.

Jak poznamenává Oudeyer (2006, s. 68–74), toto modelování se nejprve hojně rozšířilo v biologii, ale v poslední době se začíná etablovat i v lingvistice. Zároveň dodává, že jeho smyslem není vytvořit simulaci, která by co nejpriléhavěji odpovídala realitě, a tu prohlásit za jedinou možnou a správnou teorii studovaného jevu. Účelem modelů je přejít od intuitivních verbálních teorií k formálním, u nichž je možné kontrolovaně testovat vnitřní konzistenci a dopady, a obrábět tak prostor potenciálně přijatelných teorií o daném jevu.

³V aktuálním výzkumu se této problematice věnují např. Brown (2000), Forde Thompson et al. (2012).

⁴Stále se vede debata o tom, zda jsou jazyk a hudba z evolučního hlediska projevy adaptace, exaptace (využití prostředků původně vyvinutých v jiném kontextu pro nové účely), či jen vedlejším produktem jiných adaptivních procesů, jak v případě hudby tvrdí Pinker (2005, s. 20).

Napoví nám tak, které teorie má smysl dále empiricky zkoumat a které již z hlediska své vlastní logické struktury trpí nedostatky. Modely tedy mohou vycházet i z předpokladů, o nichž bezpečně víme, že jsou nesprávné – buď proto, aby se studovaný problém zjednodušil, nebo proto, aby se rigorózně vyvodilo, k jakým důsledkům ony předpoklady vedou. Zbývá zmínit, že dosavadní studie se věnují vzniku řečových i hudebních dovedností a konvencí odděleně, a není nám známo, že by se někdo pokusil o komplexní formalizaci dynamiky vztahů mezi oběma oblastmi.

Nejvíce prostoru pro empirické ověření svých hypotéz mají badatelé, kteří uvažují o **neurální integrovanosti** obou systémů, tedy o tom, že lidský mozek při produkci a percepci řeči i hudby využívá podobných strategií a neurálních cest. V centru zájmu je otázka, zda existují společné mechanismy např. pro zpracování signálu na identifikovatelné vjemové jednotky (diskretizace a klasifikace) či vytváření a organizaci mentálních reprezentací, v nichž jsou tyto jednotky uloženy (Patel 2008, s. 3–4). Psycholog Aniruddh D. Patel se touto problematikou dlouhodobě zabývá a vyslovil některé domněnky, které stojí za zamyšlení. Patří mezi ně:

- hypotéza o sdílených (myšleno hudbou a jazykem) neurálních prostředcích pro syntaktickou integraci (Shared Syntactic Integration Resource Hypothesis, **SSIRH**; Patel 2003)
- hypotéza o sdíleném mechanismu pro osvojování zvukových kategorií (Shared Sound Category Learning Mechanism Hypothesis, **SSCLMH**; Patel 2008, s. 72)

Obě již našly jistou podporu ve výsledcích percepčních i neurologických studií s použitím měřících a zobrazovacích metod jako jsou evokované potenciály (ERP) či funkční magnetická rezonance (fMRI; blíže viz oddíl 2.2.1). Tradičně se přitom soudilo, že hudba a řeč spolu kognitivně příliš nesouvisí. Tato tvrzení byla dokládána příkladovými studiemi mozkových lézí, které postihly jen jednu z obou schopností: afázie (ztráta schopnosti produkovat či vnímat řeč nebo některý z jejích aspektů) není nutně provázána amúzií (obdobné postižení vzhledem k hudebním schopnostem), a obráceně. Ovšem sám Patel tvrdí, že tato pozorování nejsou nutně nekompatibilní s existencí sdílených *mechanismů*; vypovídají spíš o tom, že sdíleny nejsou mentální *reprezentace* hudebních a řečových jednotek, a tak mohou být porušeny jedny a ne druhé, což ale není překvapivé (Patel 2008, s. 73). Konečně, jestliže oba systémy sdílejí některé netriviální mozkové struktury a procesy, nepřímo to zpětně poukazuje i na jejich možné evoluční propojení.

V naší práci se vynasnažíme nesklouzávat k metaforickým interpretacím vztahu mezi řečí a hudbou. Ač jsou mnohdy podnětné či esteticky povznášející a leccos vypovídají o kulturním prizmatu, kterým nahlížíme na svět, jejich bezbřehost by sváděla k ukvapeným (možná příliš optimistickým) závěrům ohledně mnohem prozaičtějšího, střízlivějšího a přesněji ohraničeného tématu této studie. To je blíže druhému (fyzickému) pólu škály

pohledů na vztah mezi jazykem a hudbou. Náš výzkum není přímo biologický, tj. nehledáme např. sdílené mozkové struktury aktivované jak zpracováním řeči, tak hudby (viz níže), ale vycházíme z předpokladu, že pokud takovéto sdílené struktury či mechanismy existují, měly by se odrazit ve způsobu, jakým jedinci reagují na řečové a hudební podněty. Než se ale vydáme dál, musíme si ujasnit, co přesně myslíme oněmi dvěma základními pojmy, s nimiž jsme dosud operovali poměrně volně. Co je tedy přesně hudba, zejména pokud jde o zvukovou strukturu, jež bude v centru našeho zájmu? A v čem se po této akustické stránce liší od jazyka?⁵

1.1 Zvuková charakteristika hudby

Koncept hudby zastřešuje mnoho věcí, pro něž bychom jen stěží hledali společné jmenovatele, snad kromě média (zvuk) a estetického prožitku. Zvláště formální experimenty ve 20. stol. značně rozvolnily pohled na to, co vůbec hudbou je. Slavná skladba Johna Cage 4'33" poskytne posluchačům 4 minuty a 33 vteřin ticha zahraného na libovolný počet jakýchkoli nástrojů a doplněného o zvuky prostředí, a po příchodu počítačů a elektronické hudby už je to, co je možné složit, omezeno jen naší představivostí. Nás ale zajímá hudba v tradičnějším pojetí, tedy ne jako formální experiment, nýbrž jako umělecký artefakt, který vzešel z komunitního rituálu a prožitku s významnou sociální rolí, podobně jako jazyk.⁶

S touto užší definicí hudby již nějaké obecné poznatky vyvodit lze, např. co se využití jednotlivých určujících vlastností zvuku v hudbě týče. Tyto dimenze má zvuk čtyři (až pět): výšku, délku, hlasitost, tón (tj. charakteristika, dle níž rozlišíme dva zvuky, které jsou jinak podle předchozích tří dimenzí totožné) a případně ještě směrové určení. Jejich akustickými koreláty jsou základní frekvence (F_0), trvání, hladina zvuku a spektrum (frekvenční složení, tj. poměrné zastoupení energie v různých frekvenčních pásmech) zvuku. Hudební systémy jsou nejčastěji organizovány okolo kontrastů výšky. Základní jednotky, z nichž se hudba skládá, jsou od sebe primárně odlišeny různými základními frekvencemi a mluvíme tedy o „tónovém“ materiálu jisté hudební tradice. Není přitom potřeba, aby daná kultura měla explicitní koncept stupnice, tedy jistého kanonického seřazení těchto jednotek od nejnižší po nejvyšší (Patel 2008, s. 17).⁷ I tón ovšem hraje důležitou roli,

⁵Následující dva oddíly v mnohém čerpají z pojetí a výkladu v Patelově monografii *Music, Language, and the Brain* (2008).

⁶Z filozofického hlediska by se dalo argumentovat, že takovou roli pro dnešní společnost plní právě i formální experiment, který je důležitým prostředkem kritické reflexe stávajícího řádu věcí. Princip tkví v tom, že něco, co bychom konvenčně za hudbu nepovažovali, jak tak cíleně označeno, což vede ke zpochybnění jistot. Nás ovšem zajímají charakteristiky, které právě tyto jistoty, konvence či očekávání definují.

⁷Patel dokonce zmiňuje, že opozice „vysoký-nízký“ není zdaleka jediná metafora, kterou různá etnika frekvenční rozložení tónů v hudbě popisují: američtí Indiáni Havasupai používají rozdíl „tvrdý-měkký“ a

běžně sice spíš pro dosažení širší palety zvuků jako v západní orchestrální hudbě, ale ve výjimečných případech dokonce nahrazuje výšku v roli organizačního principu, na němž je vybudován hudební zvukový materiál. Takovým případem je např. severoindický perkusní nástroj tabla (Patel 2008, s. 34–7).

Temporální organizace hudby – kontrasty trvání jejích jednotek, jinými slovy rytmus – bývá převážně založena na izochronii (periodicitě): trvání tónů je definováno v celočíselných zlomcích nějakého základního trvání (noty celé, půlové, čtvrtové atd.) a percepční dojem taktu (či pulzu s ním spojeným) je navozen pravidelným střídáním těžkých a lehkých dob. I zde ovšem existují výjimky: např. v západní hudbě dává v partu pokyn (*tempo*) *rubato* interpretovi volnost, aby volil temporální parametry produkce skladby podle svého vlastního uvážení a výrazového záměru. Je pravda, že nejčastěji tomu tak je u sólových pasáží, kde není důležitá synchronizační role jinak převládající izochronie. Ovšem v jávském hudebním žánru známém jako *pathetan* se izochronní rytmus systematicky nevyužívá navzdory tomu, že obnáší současnou produkci vícero hudebníků; ti dosahují koordinace díky tomu, že pečlivě sledují vedoucího hráče (hlavní hlas). Takové případy jsou však v hudbě spíš ojedinělé (Patel 2008, s. 100).

1.2 Zvuková charakteristika řeči

Hned zkraje zmiňme, že pro tuto práci není příliš relevantní strukturalistická dichotomie mezi jazykem jakožto abstraktním systémem a řečí jako jeho projevy, zavedená Ferdinandem de Saussurem.⁸ Budeme tedy obou pojmů využívat víceméně zaměnitelně a v případě potřeby je desambiguovat pomocí termínů jako je právě „systém“ nebo naopak „promluva“. Důležité je, že po většinu času pro nás budou jazyk i řeč odkazovat na zvukovou stránku lidské komunikace. Tu lze popsat ze stejných úhlů pohledu, které jsme před chvílí zdůraznili u hudby – z hlediska jejího spektrálního, výškového a temporálního uspořádání.⁹

Postupy využití fyzikálních vlastností zvuku pro potřeby řeči se do značné míry liší od těch hudebních. Zaprvé, převažuje témbrová organizace zvukových jednotek (Patel 2008, s. 9): všechny jazyky disponují nějakým hláskovým inventářem, jehož elementy jsou od sebe primárně odlišeny na základě spektrálních kontrastů. Nejlépe je to vidět na systémech vokálů, k jejichž identifikaci slouží zejména první dvě pásma nad základní frekvencí, která

liberijský kmen Kpelle zase „malý-velký“.

⁸Saussure přísně vzato původně postuloval trichotomii mezi jazykem (*langue*), řečí (*parole*) a pojmem *langage*, který označuje obecnou lidskou predispozici a schopnost vytvářet složité komunikační systémy (tedy jazyky).

⁹Intenzita je zvláště z hlediska percepce ošemetná a pro nás nepříliš relevantní, takže se jí blíže věnovat nebudeme. Snad jen uvedme, že intenzitní rozdíly napomáhají identifikaci slabiky: slabičné jádro je typicky místem maximální apertury, tj. otevřenosti vokálního traktu, a tedy i nejvyšší intenzity v průběhu slabiky.

ve spektru vykazují silnou koncentraci energie (tzv. první a druhý formant). Zadrugé, kontrasty výšky jsou sice bohatě využívány v prozodii, ale není úplně jasné, zda v nich vládne tak pevná úroňová kategorizace jako v hudebních tonálních systémech. Na jednu stranu máme tónové jazyky, v nichž jsou výškové difference lexikálně kontrastivní, a tudíž relativně pevné, byť experimenty, které se snaží dokázat, že tónové úroňe jsou u mluvčích v delším časovém měřítku stabilní a absolutně dané, jsou silně kontroverzní (Patel 2008, s. 47). Také byly vyvinuty teoretické modely intonace (např. autosegmentální teorie a systém ToBI), které na hloubkové úrovni postulují reprezentaci intonace jako sekvenci diskretních tónových událostí. Na druhou stranu existují plodné přístupy k intonaci, které ji interpretují jako spojitý jev, což obhajují její roli ve vyjadřování afektivních stavů – jejich proměny a intenzita jsou také vnímány jako spojitá proměnná (Patel 2008, s. 205–6).

Co se temporálního uspořádání řeči týče, najít v řečovém rytmu alespoň tendenci k podobné izochronii, jakou známe z hudby, se nepodařilo. Pokusů přitom bylo mnoho; podnět k nim zaval Kenneth Pike (1945) svým rozdělením jazyků podle rytmu na slabičně izochronní (tj. jednotlivé slabiky v promluvě jsou odděleny přibližně stejným časovým intervalem) a přízvukově izochronní (stejný časový interval v promluvě odděluje naopak přízvuky). Při nejlepší vůli se tuto hypotézu potvrdit nepodařilo a výzkum rytmu v řeči se tedy odklonil k jinému typu pravidelností a temporálních vzorců, které se na izochronii nezakládají (viz níže). Je tedy důležité zdůraznit, že navzdory tomu, co nám může napovídat zkušenost s hudbou, se rytmus nerovná nutně izochronii, a u řeči to platí dvojnásob: jde čistě o vzorce výskytu relevantních jednotek v čase a jejich vzájemné vztahy. Co mají naopak rytmus v jazyce a hudbě společného je hierarchizace prominencí: důrazy a hranice mezi jednotkami spadají do několika úrovní nápadnosti (Patel 2008, s. 140).

Je důležité podotknout, že se řeč a hudba mohou lišit nejen využitím jednotlivých zvukových dimenzí, ale i tím, do jaké míry jsou na jejich variace v obou doménách lidé jakožto posluchači citliví, resp. jak velký rozdíl je pro ně již percepčně postřehnutelný a relevantní. Patel (2008, s. 143–4) zmiňuje výzkum Ilse Lehisté, jež se snažila odhadnout práh vnímatelného rozdílu (*just noticeable difference*, **JND**) pro mezipřízvukové intervaly v řeči na základě percepčních experimentů s porovnáváním trvání zvukových vzorků šumu. Došla k závěru, že tento práh je minimálně v rozmezí 30–100 ms, pravděpodobně ale dokonce o něco výš. Je tento práh u hudby podobný? A jak vypadá srovnání, co se citlivosti na výškové kontrasty týče?

Funkčně se jazyk od hudby liší především možností přesnějšího sémantického zacílení: kromě obecných afektivních významů může být nositelem i velmi přesných věcných sdělení. V poslední době sice některé experimenty naznačují, že i hudba snad v omezené míře může nést propoziční významy (Koelsch et al. 2004), ovšem tato tvrzení stále mezi odborníky vyvolávají vášnivé půtky o přesnou interpretaci inkriminovaných pozorování (Slevc a Patel 2011, aj. ve stejném čísle časopisu).

Někdy se též uvažuje o paralele ve fonologickém aspektu řeči a hudby: v obou případech dochází k rozparcelování spojitého zvukového prostoru na diskrétní kontrastivní jednotky, které podléhají určité kombinatorice. Jak už poznamenal průkopník (nejen) experimentální akustiky Hermann von Helmholtz, soustavy tónů využívané v hudbě nejsou ničím shůry či přírodou dány a v čase se mění (Helmholtz 1885, s. 236) – podobně jako jazyky. To je sice pravda, ale třeba kategorická percepce, která je nutným předpokladem pro diskretizaci akustického kontinua a vyčlenění ustálených jednotek, je prokázána i u některých dalších obratlovců včetně čičil a ptáků, takže jde spíš o poměrně ranou adaptaci sdílenou širším spektrem obratlovců, která má usnadnit zpracování zvukového signálu (Hauser et al. 2002, s. 1572). Není tedy úplně jasné, zda řeč a hudba ve svém fonologickém rozměru sdílejí pouze tuto obecnou adaptaci, což by na žádný zvláštní vztah mezi oběma doménami neukazovalo, nebo zda spřízněnost sahá dál.

2 Spřízněné domény?

Ujasnili jsme si tedy, co pojmy řeč a hudba míníme; vraťme se nyní trochu systematictěji k jejich rozdílům a podobnostem. Celá řada jich je nasnadě – mezi podobnostmi např. to, že u obou je dominantním médiem zvuk nebo že obě jsou v mnoha ohledech „fonologickými“ systémy (Patel 2008, s. 9), neboť spojitou zvukovou matérii člení do kulturně stanovených kategorií. To jsou ovšem shody poměrně obecného rázu a při bližším ohledání se navíc začnou jevit problematickými: viděli jsme třeba, že ačkoli hudba i řeč pracují se zvukem, jeho dostupných akustických dimenzí využívají v různé míře a s lišícími se účely.

Kam vedly snahy souvislosti mezi řečí a hudbou zkoumat blíže a rigorózněji? Čím je možné podložit intuitivně zřejmé paralely, a co je naopak zpochybňuje? Podíváme se nejprve na možné evoluční propojení mezi oběma dovednostmi (odd. 2.1), a následně na jeho odraz ve fungování mozku (odd. 2.2.1). Pak se přesuneme k přehledu metod a měr, jež umožňují kvantifikovat rytmickou a melodickou variabilitu v hudbě i řeči navzájem srovnatelným způsobem (odd. 2.2.2). Taktéž se podíváme se na problematiku právě postřehnutelného rozdílu v temporální a frekvenční zvukové doméně (odd. 2.2.3), která úzce souvisí s tvorbou stimulů pro náš experiment. Na závěr této kapitoly se ještě podíváme na některé dosavadní výzkumy srovnávající percepci hudby a řeči pomocí poslechových testů, a jsou nám tak metodologicky nejbližší (odd. 2.2.4).

2.1 Evoluční pohled

Jak jsme již zmínili, z kognitivně evolučního hlediska jsou řeč i hudba projevy komplexního adaptivního chování (*complex adaptive behavior*), a tudíž jsou jejich vývoj a osvojování silně podmíněny kulturními faktory, tedy podmínkami, v nichž k tomuto chování dochází. U hudby je dominantní pozice kulturní evoluce jaksí nasnadě; ne tolik už u jazyka, jehož vědecké uchopení bylo v druhé půlce 20. stol. silně formováno Chomského pojetím generativní gramatiky a jeho důrazem na genetické ukotvení jazyka. Přitom se jeví, že právě kulturní faktory hrají v ontogenezi a fylogenezi řeči rozhodující roli a úloha vrozeného je mnohem menší, než se myslelo (Steels 2011). V tomto ohledu by si tedy hudba a řeč měly být hodně blízko, při osvojování by měly začínat z podobných startovních pozic, a interakce tudíž zdaleka nejsou vyloučeny. Skutečně se např. zdá, že pokud se dvě kultury liší rytmem v řeči, liší se i převažující rytmické vzorce v jejich hudebních tradicích (Patel a Daniele 2003; Patel et al. 2006, viz 2.2.2); nedávné percepční pokusy se pak snaží ukázat, že posluchači jsou si těchto rozdílů vědomi (Hannon 2009, viz 2.2.4).

Když mluvíme o evoluci jazyka, nelze nezmínit nesmírně vlivný byť kontroverzní článek Hausera, Chomského a Fitcha „The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?“ (volně: „Jazyk jako dovednost: v čem spočívá, kdo jí disponuje a jak se vyvinula?“). V něm se v roce 2002 autoři pokusili obnovit zájem o interdisciplinární

bádání nad evolucí jazyka, spojující evoluční biology, neurovědce, psychology a lingvisty, a nastínit, kterými směry by se takový výzkum mohl ubírat. Zavádějí pojmy pro dva různé pohledy na to, co patří mezi konstitutivní rysy schopnosti jazykové komunikace (Hauser et al. 2002, s. 1571):

- na jedné straně úzce definovaná **schopnost vytvářet rekurzivní struktury** (např. že každou větu lze zapustit do konstrukce typu „Řekla, že...“, teoreticky opakovaně *ad infinitum* – „Řekla, že řekla, že řekla...“; to je schopnost jazyka v úzkém slova smyslu (*faculty of language in the narrow sense*, **FLN**)
- na straně druhé, šířeji pojaté spektrum vlastností, které nám řečovou komunikaci umožňují: kromě **rekurze** tento výčet zahrnuje i **senzoricko-motorické** a **konceptuálně-intencionální** systémy; to je schopnost jazyka v širším slova smyslu (*faculty of language in the broad sense*, **FLB**)

Těžiště článku spočívá v mapování prostoru hypotéz, které lze o evoluci těchto dvou dovedností a jejich subsystémů vytvářet a ověřovat. Tento prostor zahrnuje tři dimenze:

1. Je daná schopnost vlastní pouze lidem, nebo ji sdílíme i s dalšími živočichy?
2. Probíhal její vývoj postupně nebo skokově?
3. Vznikla přímo jako adaptace pro účely komunikace, nebo se vyvinula v jiném kontextu a posléze byla pro komunikaci pouze „rekrutována“ (tzv. exaptována)?

Autoři se domnívají, že zatímco mnohé aspekty FLB lidé pravděpodobně sdílejí s dalšími tvory, FLN je jim ryze vlastní. Zároveň ale jedním dechem zdůrazňují, že je to pouze hypotéza, kterou je třeba ověřit či vyvrátit pomocí empirického komparativního (myšleno mezidruhového) biologického výzkumu (Hauser et al. 2002, s. 1578). Dokonce se domnívají, že ačkoli byla v minulosti hlavně aplikována na primáty, tj. blízké příbuzné lidí, má smysl ji rozšířit i na další druhy obratlovců. Jen tak bude možné zjistit, zda např. právě rekurze náhodou nevznikla původně ne v kontextu komunikace, ale třeba pro usnadnění orientace v teritoriu či složitých sociálních vztazích.

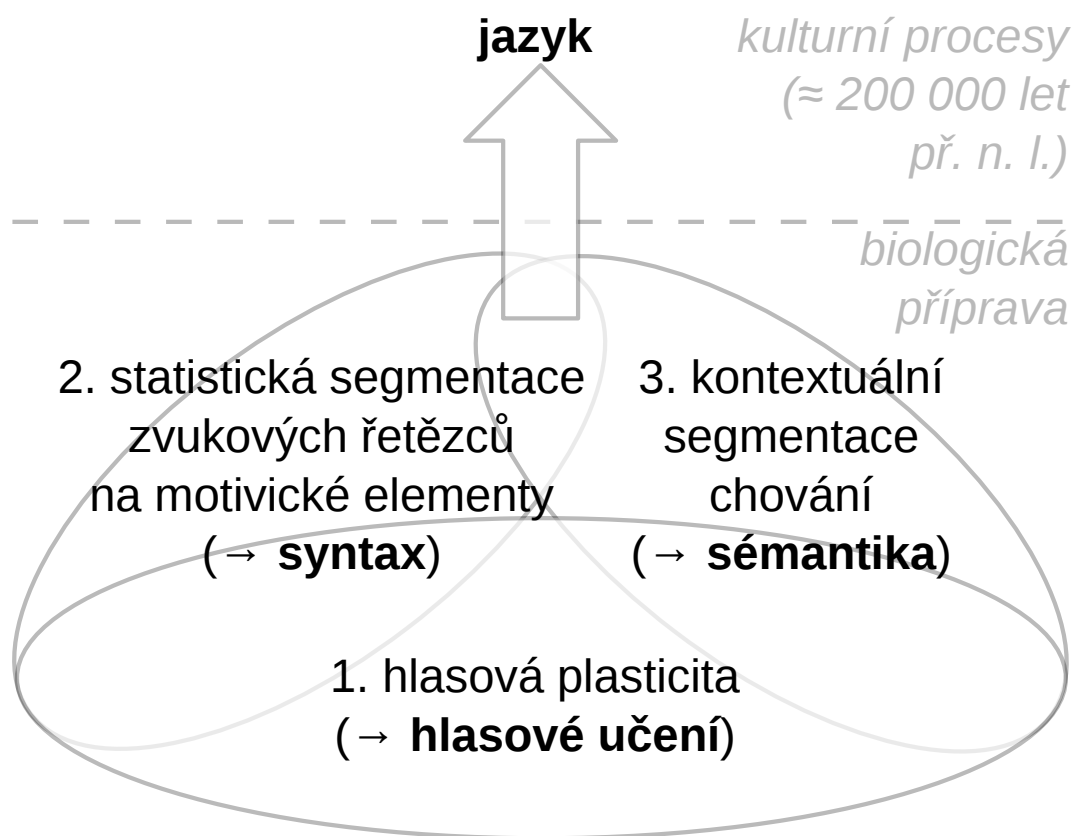
Na tuto výzvu se snaží reagovat neuroetolog Kazuo Okanoya. Zkoumá komunikaci zpěvného ptactva a hlodavců, a pomocí behaviorálních experimentů a uměle vyvolaných mozkových lézí se pokouší přijít na to, které mozkové struktury se na komunikaci podílejí a jakou při tom hrají roli.¹⁰ Jeho tezí je, že toho, co s těmito druhy sdílíme, je víc, než by se na první pohled mohlo zdát. Taktéž navrhuje trochu komplexnější pojetí jedné

¹⁰V členění následujícího výkladu budeme vycházet z aktuálního Okanoyova pojetí těchto tezí, tak jak je přednesl ve svém příspěvku „Bird song as a musical protolanguage“ („Ptačí zpěv jako hudební protojazyk“) na letní škole *International Summer School on Agent-based Models of Creativity* (Okanoya 2013); video z přednášky je dostupné na <http://www.youtube.com/watch?v=NlfVeAE1Eo8>. Paralelně budeme odkazovat i na články, kde je možné tato tvrzení dohledat v písemné podobě.

z dichotomií, kterými Hauser, Chomsky a Fitch nahlíží na prostor možných teorií o evoluci jazyka (jde o bod č. 2 výše uvedeného seznamu). Pro Okanoyu neprobíhal vývoj jazykové schopnosti přísně vzato ani skokově, ani postupně: spíš nejprve došlo k několika důležitým preadaptacím (senzomotorický aparát, konceptualizační dovednosti – **FLB**), načež samotný jazyk (**FLN**) vznikl díky interakci těchto systémů jako jejich **emergentní vlastnost** (Okanoya 2007, s. 272–3). Podívejme se blíže, jak si tyto systémy a jejich interakce představuje.

Okanoyovo pojetí je shrnuto v obr. 1. Jak vidno, zahrnuje tři hlavní subsystémy či biologické preadaptace, které samy o sobě schopnost řeči nijak nezaručují, ale z jejichž součinnosti a souhry může řeč vzejít. Od chvíle, kdy se tak stane, přechází celý proces do oblasti kulturní evoluce, čímž se oproti evoluci biologické mnohonásobně urychluje. První důležitou preadaptací je **vokální plasticita**, která vede ke schopnosti **hlasového učení** či **imitace**. Zdaleka ne všechna zvířata, která se dorozumívají zvukovými signály, se je totiž učí ze svého okolí: např. vokalizace všech primátů kromě člověka jsou čistě pudové (Okanoya et al. 2007, s. 363). Je nasnadě, že bez schopnosti vokální imitaci si nemůže jedinec osvojit zvukový repertoár, který není vrozený, ale vyvinul se na základě procesů kulturní evoluce, a tudíž se odvíjí nejen od příslušnosti k druhu, ale i ke konkrétnímu společenství. Druhou důležitou neurální oporou řeči je **statistická segmentace zvukových řetězců**, která nám umožňuje vyčlenit ze zvukového kontinua opakující se sekvence, identifikovat je jako diskrétní jednotky a dovítit se, že jednotlivé konkrétní realizace odkazují ke stejné abstraktní entitě. Je tedy předpokladem pro následné kombinatorické užití těchto jednotek, tj. **syntax**. Posledním členem základní trojice preadaptací je **segmentace kontextu**, v němž k různým typům zvukových projevů u členů druhového společenství dochází. Tato schopnost umožňuje asociovat s jednotlivými vokalizacemi různé významy a je tím pádem nezbytná pro **sémantickou stránku** jazyka.

Ovšem velmi podobné principy a požadavky hrají roli jak při zpracování hudebních stimulů, tak při vlastním hudebním projevu, a to nejen u lidí, ale právě např. u zpěvných ptáků. Z hlediska percepce je třeba zpracovat řetězec jemných akustických variací v čase a nasegmentovat ho na smysluplné kousky, které lze posléze použít ve vlastní produkci. Ta je přitom alespoň ve své vokální podobě umožněna právě schopností přesně ovládat motoriku vokálního traktu. Jsme si vědomi toho, že termín „zpěv“ pro ptačí zvukovou komunikaci je spíše metaforou, a tudíž se jedná o volnou paralelu s hudbou v lidském slova smyslu, ne o přímou analogii. Jde ovšem o paralelu plně zamýšlenou a cílenou: jejím smyslem je podložit hypotézu, že lidé podstatnou část aparátu, který využívají pro jazykovou komunikaci, sdílejí s jinými druhy a že dokonce existuje kontinuita (jak synchronní, tedy mezidruhovává, tak diachronní, tj. v rámci vývoje lidského druhu) mezi zvířecími zvukovými signály, jako je např. zpěv ptáků či velryb, a lidskou řečí. Z tohoto pohledu se tedy hudba a jazyk zpočátku vyvíjely společně (jakýsi hudební či zpěvný



Obrázek 1: Nástin evoluce jazyka. Převzato z Okanoyi (2013).

protojazyk) a teprve později se disociovaly. V některých případech dokonce ne úplně, jak tomu nasvědčuje důležitá společensko-organizační a komunikační funkce hudby, jež dodnes přetrvává¹¹ v jistých společenstvích, jako je v úvodu zmíněný pygmejský kmen Mbendjele. Proč se tyto predispozice vyvinuly jen u některých živočišných druhů, a proč právě u nich? A do jaké míry jsou spřízněné i jejich neurologické substráty?

Schopnost vokální imitace je kromě lidí vlastní např. netopýrům, velrybám a některým druhům ptáků (Okanoya et al. 2007, s. 366). Z neurálního hlediska pravděpodobně souvisí s přímým propojením mezi orofaciálními motorickými oblastmi mozkové kůry a strukturami v prodloužené míše, které mají na starost dýchání: přítomnost této nervové dráhy odlišuje lidi od ostatních primátů a zpěvné ptáky od nezpěvných (Okanoya et al. 2007, s. 363 a jeho obr. 1). Touto drahou se uzavírá smyčka produkce a percepce, která umožňuje imitaci: původně autonomní (dýchací) a emocionálně-pudové (vokalizační) systémy se dostávají napojením na mozkovou kůru pod vědomou kontrolu a za předpokladu, že na vstupu existuje jakýsi zrcadlový systém, který usouvztažňuje vjemy zpracovávané auditorní oblastí kůry s odpovídajícími artikulační úkony, může dojít k postupně vylepšovaným pokusům o reprodukci slyšeného (Okanoya 2007, s. 274–5 a jeho obr. 4). Bez tohoto

¹¹Když říkáme „dodnes přetrvává“, implicitně tím hypotézu o hudebním protojazyku v nějaké formě přijímáme. Pokud se prokáže jako nesprávná, bude pochopitelně namísto říkat, že se tato role hudby v některých společenstvích „nově vyvinula“.

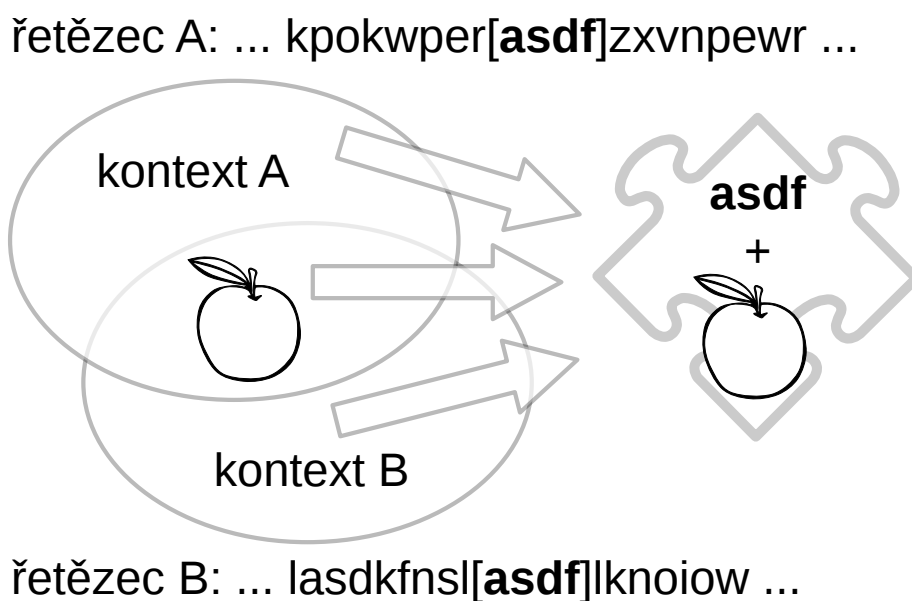
propojení se jistá volní kontrola nad hlasovou produkcí vztahuje pouze na rozhodnutí, zda vokalizaci provést či ne; ne už tak ne její konkrétní artikulační podobu.

Ať už toto spojení u živočichů schopných hlasového učení vypadá v konkrétních detailech jakkoli, v nějaké podobě přítomno být musí a je namístě se ptát, jaká mohla být motivace k jejímu vývoji. U netopýrů je důvod této adaptace nasnadě vzhledem k tomu, jak moc se místo zraku spoléhají na sluch. U velryb se domníváme, že se jemné ovládání dechu vyvinulo ve chvíli, kdy jejich prapředci začali žít v moři a museli se naučit dech zadržovat; uvědomme si, že suchozemští savci zadržovat dech běžně nedokážou (Okanoya 2013). U lidí a ptáků (konkrétněji zpěvných ptáků a papoušků) je už důvod této adaptace ošemetnější. Co se ptáků týče, uvažuje se o možnosti, že by nějak souvisela se schopností létat, která vyžaduje přesné ovládání dechového svalstva (Okanoya et al. 2007, s. 366); pak by ovšem bylo zvláštní, že se hlasové učení nevyskytuje napříč ptačí říší, ale jen u vybraných druhů. Pokud jde o lidi, jedna z odvážnějších hypotéz o vývoji lidského druhu tvrdí, že jeden z relativně nedávných článků naší evoluční linie pobýval ve vodě – tzv. Homo-Aquarius – a motivace pro rozvoj vědomého ovládání vokálního traktu by tedy byla podobná jako u velryb (Okanoya et al. 2007, s. 366). Pro toto pojetí lidské evoluce ovšem neexistují žádné fosilní doklady, a navíc by takový předchůdce z logiky evoluční dynamiky (pomalé hromadění drobných změn) musel být i předchůdcem dnešních blízkých příbuzných člověka (primátů), kteří přitom schopností hlasové imitace nedisponují.

Okanoya et al. navrhuje hypotézu, která se zaštiťuje dětským pláčem jakožto faktorem, který tento vývoj odstartoval (2007, s. 366). Pro většinu zvířat je pláč z hlediska evoluce nevýhodou, pouze upozorní predátory na snadnou kořist. Musel se tedy plně rozvinout až ve chvíli, kdy lidské společenství bylo natolik vyspělé a sociálně organizované, že si dokázalo zajistit rozumnou míru bezpečí, a větší hrozbou byla nedbalost ze strany pečovatele. Na rozdíl od ostatních primátů, jejichž mláďata mohou doslova viset na srsti svých matek a nemají tedy důvod dožadovat se pozornosti křikem, musela být nejspíš už tehdy lidská miminka pravidelně odkládána a potřebovala tedy způsob, jak se svými pečovateli manipulovat (Okanoya 2013). Podrobné studium dětského křiku ukazuje, že zatímco rané vokalizace jsou pravděpodobně zcela pudové, celkem záhy (dva týdny po porodu) začnou nabývat na složitosti, a to do té míry, že matky poznají, co konkrétně jejich potomkovi chybí (Okanoya et al. 2007, s. 367). Co je však nejpozoruhodnější, podobné vzorce chování nalezneme u druhů ptáků, jejichž mláďata jsou altriciální neboli nidikolní, tj. závislá na rodičích. Hlasité štěbetání, kterým holátka žadoní o potravu, také potenciálně může přilákat predátory, a nutí tak rodiče, aby své potomky utišili zvýšenou pozorností a péčí (Okanoya et al. 2007, s. 366).

Věnujme se nyní ještě chvíli stručně segmentaci zvukových řetězců a kontextu, dvěma zbývajícím preadaptacím z obr. 1. Tyto procesy nejspíš probíhají paralelně, resp. vzájemně, a mechanismus je podobný, ať se jedná o prvotní vývoj hudebního protojazyka či

osvojování již existující řeči. Jak naznačuje obr. 2, posluchač v proudu vokalizací identifikuje opakujících se útržky, přičemž zároveň věnuje pozornost kontextu. Když se napříč různými interakcemi částečně překrývají jak kontexty v některých svých aspektech, tak vokalizace v některých svých motivech (podřetězcích), může z toho posluchač vyvodit asociaci mezi zvukovým projevem a situací, v níž byl pronesen, např. „tento typ vokalizace či melodický motiv je spojen s příležitostí, kdy si chceme navzájem dodat odvahy; tento zase se svoláváním k jídlu“ apod.



Obrázek 2: Vzájemná segmentace zvukových řetězců a kontextu; popis viz text. Podle Okanoyi (2007, s. 273 a jeho obr. 3).

Výzkum na pěnkavách druhu *Lonchura striata domestica*, který postupně vznikl asi před 250 lety procesem domestikace, poukazuje na možné neurální substráty statistické segmentace zvukových řetězců. Zpěv jejich divokých druhů *Lonchura striata* je poměrně jednotvárný, což pravděpodobně souvisí s tlakem, který je v prostředí plném podobných konkurenčních druhů vyvíjen na vzájemnou identifikaci mezi příslušníky téhož druhu (Kagawa et al. 2012). Naopak pro domestikované pěnkavy je příznačný bohatý repertoár, který si v mládí osvojují tím, že „vyzobávají“ motivy ze zpěvu starší generace skládají z nich své vlastní projevy. Probíhá u nich tedy právě výše zmíněná statistická segmentace, tj. identifikace jednotek na základě opakování, a jejich následná recyklace (Takahashi et al. 2010, s. 481). Syntax těchto písní je poměrně složitá, zahrnuje mnoho možností kombinace a opakování jednotlivých motivů a formálně ji lze vyjádřit konečným automatem, viz např. obr. 3 v (Berwick et al. 2011). Z neurálního hlediska se pro analýzu a syntézu takto syntakticky složitých struktur ukazují jako klíčové některé korové útvary (prefrontální kortex) a bazální ganglia. Uměle přivoděná léze prefrontálního kortexu totiž nápadně výrazně zjednoduší, což lze přímočaře kvantifikovat na počtu stavů a přechodových hran

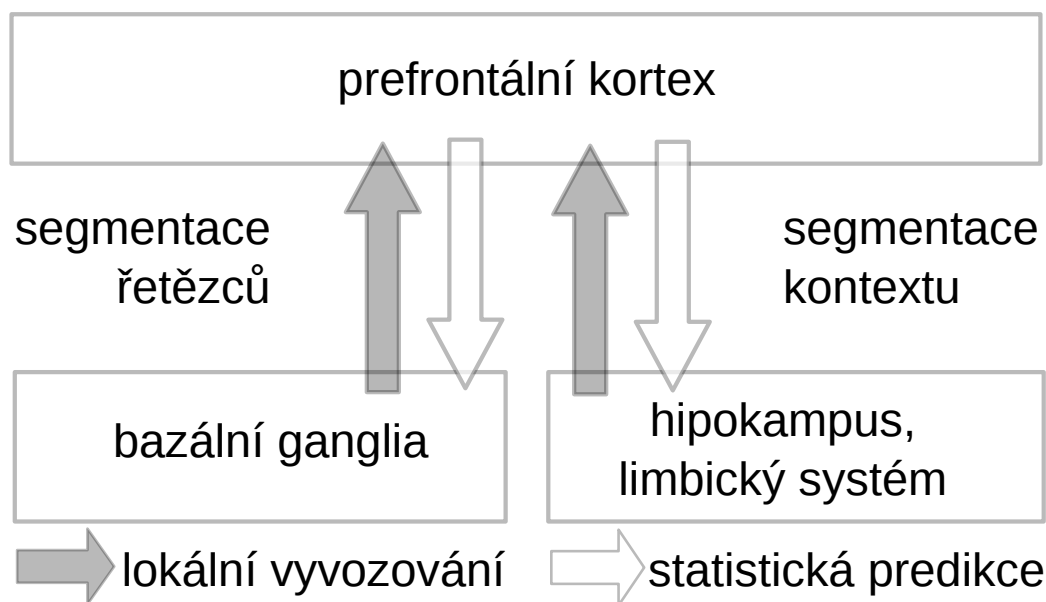
konečného automatu, který ji popisuje (Okanoya 2004). Postižena je tedy syntaktická a lexikální (ve smyslu dílčích motivů) bohatost projevu, což odpovídá vyšším plánovacím funkcím mozkové kůry. Léze bazálních ganglií pak postihuje produkci na nižší úrovni tím, že navozuje zadrhávání (Kobayashi et al. 2001).

Jak ve svém přehledovém článku popisuje Berwick et al. (2011, s. 119–20), psychologická realita a celistvost jednotlivých úseků (motivů) z hlediska reprezentace v ptačím mozku byla testována podobnými metodami, jako psychologická realita jazykových jednotek v lidské mysli. Výzkumná metoda, kterou zavedli Fodor a Bever (1965), spočívá ve zkoumání sluchové percepce sekvencí slov s uměle vloženými lupnutími, přičemž posluchači mají identifikovat, kdy lupnutí nastalo. Pokud je oproti skutečnému okamžiku někam posunou, interpretuje se ono místo jako hranice jednotky, kterou mozek považuje za celek (např. slabika, slovo, přízvukový takt). Stejný typ experimentu, ovšem provedený na domestikovaných pěnkavách a s použitím tří- až čtyřtónových sekvencí místo jazykových jednotek, přinesl podobné výsledky: taktéž se u nich projevuje tendence přemísťovat percepční dojem lupnutí na hranice motivických celků (Suge a Okanoya 2010), což by naznačovalo, že jsou v jejich mozku reprezentovány jako ucelené jednotky. Z jiného pohledu psychologickou realitu těchto jednotek potvrzují pokusy využívající měření evokovaných potenciálů u lidí (**ERP**; jde o výkyvy napětí na elektrodách umístěných na různých částech hlavy, vázané na percepci nějakého vjemu). Bylo prokázáno, že lidé jsou schopni z plynulého toku stejně dlouhých not po nějaké době vydělit třítónové sekvence jako celky, pokud se tyto sekvence v jinak nediferencovaném sledu tónů opakují (Abla et al. 2008, s. 963). Na úrovni evokovaných potenciálů se proces seskupování tónů do větších celků na základě statistických pravidelností projevuje výrazným negativním komponentem na křivce napětí, který se objeví přibližně 400 ms po začátku prvního tónu z třítónového celku.¹² Zbylé dva tóny už takovou reakci nezpůsobují, z čehož se právě usuzuje, že v percepci tvoří jednotku dohromady s oním prvním. Studie se srovnatelnými výsledky byly provedeny i na předem neznámých jazykových jednotkách, tj. pseudoslovech (viz přehled v Abla et al. (2008), 953). Sluchová percepce hudebních i řečových řetězců a orientace v nich se tedy u lidí zakládá na stejných principech statistické segmentace. To samo o sobě není až tak překvapivé – statistická segmentace je robustní mechanismus a jako taková je hojně využívána napříč druhy (některé opice) i doménami (zrak – obojí viz Abla et al. 2008, s. 962). Zajímavější je, že podle výzkumu, který v současné době Dilshat Abla a další provádějí, se zdá, že v procesu této aktivní (*on-line*) segmentace se angažují právě prefrontální kortex a bazální ganglia (Okanoya 2013), tedy oblasti, jejichž léze u pěnkav způsobuje komplikace při produkci zpěvu a její zjednodušení (viz výše).

Komparativní neuroetologický výzkum segmentace kontextu je již složitější, zejména proto, že zvířecí komunikační kódy běžně neoplývají vysokou měrou sémantické sofisti-

¹²V literatuře o ERP je technické označení pro takový komponent N400.

kovanosti, a je tedy těžší určit, zda k segmentaci kontextu (přiřazení významu) dochází správně či ne. Nicméně pokud si segmentaci kontextu vyložíme jako schopnost správně jej interpretovat a volit podle něj vhodné vzorce chování odpovídající danému druhu, pak lze pomocí stejného experimentálního paradigmatu jako výše vyvodit, že u např. u pěnkav a osmáků hrají v této schopnosti důležitou roli amygdala, resp. hipokampus. Léze amygdaly u pěnkav vede k rozkladu pářícího rituálu – sameček má tendenci úplně vynechat fázi dvoření (Okanoya 2013); léze hipokampu u osmáků, jinak velmi sociálně založeného druhu hlodavce, vede k agresivitě, která se projevuje tím, že na opečovávání a čištění od jiného osmáka reaguje postižený jedinec místo reciproční služby útokem (Uekita a Okanoya 2011, s. 307). Ve výsledku si Okanoya neurální implementaci výše popsaných procesů vzájemné segmentace zvukových řetězců a kontextu představuje tak, jak nastiňuje obr. 3. V nevědomých strukturách mozku, jako jsou bazální ganglia, hipokampus nebo limbický systém, probíhá základní segmentace založená na čistě lokálních kritériích. Informace jsou dále postoupeny korovým strukturám (prefrontálnímu kortexu), kde jsou integrovány jak vstupy ze segmentace kontextu (sémantika), tak vstupy ze segmentace řetězců (fyzické nosiče a jejich kombinatorika, tj. syntax). Jakmile prefrontální kortex nashromáždí dostatek informací o rozdělení těchto jednotlivých elementů, jejich souvěskytech apod., může začít dávat zpětnou vazbu nižším mozkovým strukturám a podílet se na segmentaci pomocí statistické predikce.



Obrázek 3: Neurální substráty vzájemné segmentace zvukových řetězců a kontextu; popis viz text. Podle Okanoyi (2013), ale viz též jeho výklad v (Okanoya a Merker 2006), cit. v (Okanoya 2007, s. 276).

Úhrnem: přímé experimenty s evokovanými potenciály ukazují, že principy statistické segmentace, které uplatňujeme na řetězce tónů, používáme i při identifikaci řečových

jednotek. Co víc, tuto schopnost mají již kojenci a při akvizici motivů, z nichž skládají svůj zpěv, ji využívají i zpěvní ptáci. Též víme, že v rámci ontogenetického vývoje jedince existuje jak u těchto ptáků, tak u lidí privilegované období, v němž k osvojování vokalizačních vzorců dochází nejjednodušeji (Bolhuis et al. 2010, s. 749). Ptačí způsob akustické komunikace je přitom blíže hudbě v tom ohledu, že neposkytuje příliš jemnou sémantickou diferenciaci lokalizovatelnou na jednotlivé motivy, tak jak to plně rozvinutá řeč umožňuje se slovy. Zaměřuje se spíše na redundantní znázornění silných emocionálně-pudových významů (zejm. sexuální zájem) a v tomto ohledu se tedy zdá být příbuzný s hypotetickým hudebním protojazykem, z něhož se u lidí mohla vyvinout řeč na straně jedné a hudba na straně druhé.

2.2 Výzkum temporální a frekvenční struktury řeči a hudby

Jelikož se v experimentální části práce hodláme zaměřit na percepční test založený na porovnávání řečových a hudebních stimulů, bylo by záhodno si ujasnit, v jakém smyslu a v jakých intencích jsou tyto dvě kategorie akustických vjemů z hlediska lidské percepce vůbec souměřitelné. Některé studie lidského mozku s využitím zobrazovacích metod nám umožní pohlédnout na Okanoyovy argumenty o spřízněnosti řeči a hudby (uvedené v oddíle 2.1) z jiného úhlu: zatímco Okanoyovo pojetí je diachronní a komparativně biologické, tyto studie se přímo zabývají fungováním lidských neurálních struktur, a to ze synchronního hlediska, bez ohledu na možné vývojové linie, které jejich současné nastavení zapříčinily. Dále tak upevní pracovní hypotézu, že zpracování řeči a hudby je v lidském mozku do jisté míry propojené a že toto propojení se může odrážet v korelaci percepčních dispozic jedince vůči oběma druhům zvukových vjemů.

Jak uvidíme níže, často citovaným vysvětlením pro nalezené sdílené neurální mechanismy je akustická strukturní podobnost řeči a hudby.¹³ Proti tomuto tvrzení lze ovšem postavit dva relativně závažné protiargumenty. Zaprvé, jak jsme již zmínili v oddílech 1.1 a 1.2, řeč a hudba využívají dostupné akustické dimenze v rozdílné míře a rozdílnými způsoby: u hudby převažující důraz na přesné variace výšky a trvání je v řeči potlačen (resp. rozvolněn), velkou významotvornou roli hrají naopak variace spektrálního složení (zabarvení). Zadruhé, formulace, které se snaží tuto strukturní podobnost popsat, znějí většinou velmi obecně, např. řeč a hudbu zařazují do kategorie „jemně strukturovaných stimulů, které se mění v čase“¹⁴ (Levitin a Menon 2003, s. 2142). To je ovšem spíše popis sluchové percepce jako takové než specifikum obou jmenovaných modalit.

Jackendoff jde dokonce tak daleko, že připouští bližší spříznění pouze na úrovni abs-

¹³Z pohledu Okanoyovy teorie je tato podobnost důsledkem evolučního spříznění řeči a hudby, ale i kdybychom s jeho hypotézou nesouhlasili či se prostě zdrželi soudů o evolučním aspektu věci, nic nám nebrání uvažovat o oné podobnosti čistě ze synchronního hlediska, ať už vznikla jakkoli.

¹⁴“ [...] fine-structured stimuli that evolve over time [...] ”

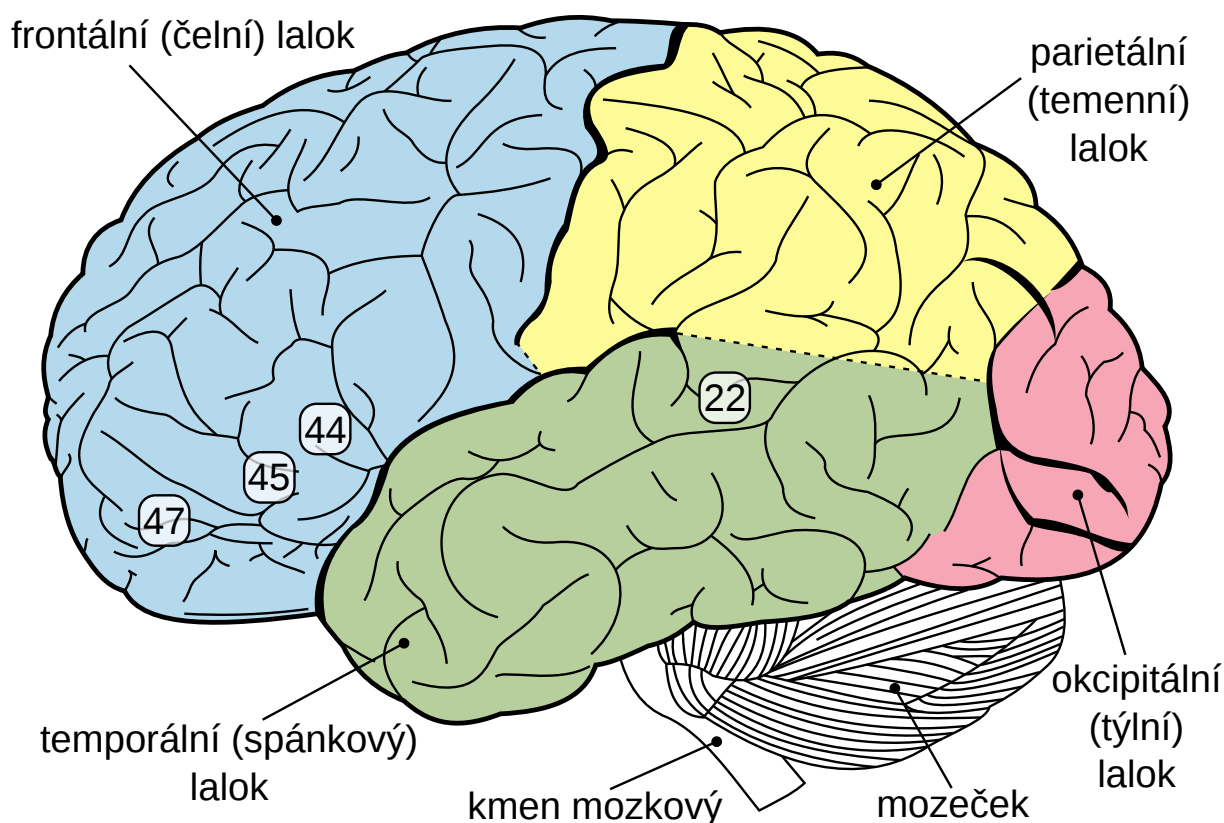
traktní metrické struktury, která řídí temporální sled slabik/tónů (2009, s. 199); zde je nejspíš pod vlivem generativní teorie tonální hudby, kterou v 80. letech vyvinul spolu s Fredem Lerdahlem. Generativistická hloubka je ovšem v celé záležitosti spíš na přítěž, předně proto, že nutně zabíhá do podrobností jazyka či hudby jako systému a upozaduje tak povrchové akustické jevy, které jsou přitom v centru našeho zájmu. Když se na celou záležitost podíváme spolu s Patelem (2006) trochu pragmatičtěji, zjistíme, že existující metody pro prozodická měření, jako jsou např. normalizovaný index párové variability (*normalized pairwise variability index*, nPVI) nebo prozogram, umožňují kvantifikovat ať rytmické či melodické podobnosti mezi hudbou a řečí s až překvapivě informativními a smysluplnými výsledky. Bude jim proto věnován oddíl 2.2.2.

2.2.1 Hudba a řeč v mozku

Studie aktivací jednotlivých oblastí mozku v závislosti na vykonávané činnosti či vnímaných podnětech jsou možné díky tzv. **funkčním zobrazovacím metodám**, přičemž slovo „funkční“ zde odkazuje ke schopnosti mozek nejen anatomicky zmapovat, ale i zjistit, na jakých místech v daném okamžiku dochází ke zvýšené aktivitě, o níž se předpokládá, že souvisí s jistou funkcí. Narážejí ovšem na mnohé teoretické i praktické překážky. Na teoretické rovině se vede spor o to, zda je představa, že jednotlivé typy činností či stimulů jsou řízeny a zpracovávány ohraničenými oblastmi (tzv. **lokalizacionismus**), vůbec validní. Z praktického hlediska musí být tyto experimenty velmi důsledně kontrolované, aby se zaručilo, že změřené aktivace skutečně souvisí s experimentálním podnětem a nejsou projevem nějakého souběžného rušivého vlivu. To se u mozku, jenž neustále i nevědomě zpracovává vjemy a stará se o chod životně důležitých funkcí, zaručuje poměrně špatně. U výzkumů zaměřených na sluchovou percepci je navíc problém, že měřicí přístroje u některých metod (např. u **funkční magnetické rezonance**, fMRI) jsou hlučné a je tedy potřeba stimuly od nich odstínit (ať už fyzicky či temporálně, tj. tím, že se přístroj spustí až po zaznění stimulu). Snaha najít v získaných datech relevantní aktivační vzorce pak občas připomíná hledání jehly v kupce sena, což s sebou nese další problém: stává se, že se najde neexistující jehla. Vědci se totiž často spokojí se statistickou významností svého výsledku, jíž přitom při dostatečně velkých datech není takový problém dosáhnout, a nevěnují již pozornost tomu, že samotná velikost zjištěného efektu může být zanedbatelná.

S vědomím těchto omezení se nyní podíváme na studii Abramse et al. (2011), v níž autoři zjistili, že pro dekodování temporální struktury jak řeči, tak hudby existují v mozku sdílené mechanismy či oblasti, jež se ovšem zároveň v některých drobnostech (v detailním prostorovém uspořádání aktivací) částečně i liší. Deklarovaným cílem článku je podpořit výše zmíněnou Patelovu hypotézu o sdílených neurálních prostředcích pro syntaktickou integraci (SSIRH), ale díky specifickému charakteru použitých stimulů, jež se liší od dřívějších výzkumů v této oblasti, je možné výsledky interpretovat i v tom smyslu, že poukazují

na roli rytmu při zpracovávání řeči i hudby (viz níže). Syntaxí je zde míněno významotvorné řazení základních stavebních jednotek daného systému a SSIRH by se dala shrnout následovně: syntax v hudbě i jazyce je zpracovávána ve stejné oblasti mozku (v prefrontálním kortexu; viz obr. 4), ale samotné jednotky, tzv. „strukturní reprezentace“, jsou uloženy jinde, v posteriorních oblastech mozku (Patel 2003, s. 674).¹⁵ Většina předchozích studií na toto téma se věnovala buď hudbě, nebo řeči, a jisté povzbuzující výsledky v tomto směru přinesla. Záměrem Abramse et al. bylo provést přísně kontrolovaný komparativní výzkum na stejných subjektech a se srovnatelnými stimuly v obou doménách (Abrams et al. 2011, s. 1507).



Obrázek 4: Topologické členění mozku na laloky (viz popisky) a **Brodmannovy oblasti** zmíněné v textu (viz číselné štítky). **Brocovo** centrum řeči tradičně odpovídá BA 44 a 45, **Wernickeovo** pak BA 22.

Výzkumný tým vyvinul skutečně úctyhodnou snahu o to, aby minimalizoval vliv parazitních proměnných v použitých nahrávkách na výsledně zjištěné aktivace. Pomocí dotazníku v pilotní studii byly řečové i hudební stimuly (slavné projevy, resp. klasická hudba)

¹⁵Přitom jak též tvrdí Patel (2008, s. 84), mechanismus *akvizice* těchto jednotek může být znovu sdílený a podle něj spočívá ve statistické generalizaci na základě slyšených realizací, při níž jedinec indukuje charakteristiky rozdělení nějakého jevu podle jednotlivých pozorování. Jde o v úvodu letmo zmíněnou hypotézu o sdíleném mechanismu pro osvojování zvukových kategorií (*Shared Sound Category Learning Mechanism Hypothesis*).

ohodnoceny z řady hledisek (emocionalita, vzrušivost, povědomost) a do ostrého testu byly následně vybrány co nejpečlivěji spárované stimuly z obou domén (Abrams et al. 2011, s. 1508–9). Z těchto vybraných syntakticky validních stimulů byly na základě metody vyvinuté Levitinem a Menonem (2003) vytvořeny stimuly syntakticky narušené.

Tato metoda stojí za bližší pohled. Dřívější studie zaměřené na neurální zpracování syntaxe se totiž zakládaly na experimentálním paradigmatu, v němž byla mezi syntakticky validní stimuly vmezeřena nahrávka, která obvyklou syntax nějak lokálně porušovala. V hudbě toto porušení syntaxe může spočívat např. ve vybočení z předem navozené tóniny/harmonie; v jazyce pak v nerespektování konkrétní instance syntaktického pravidla ve větě. Takové stimuly se ovšem od těch syntakticky správných z hlediska zpracování a aktivací v mozku neliší jen narušenou syntaxí, ale i tím, že vyvolávají překvapení, případně si vyžadují aktivní opravu mentálně budované syntaktické struktury – přeorientování na jiné tonální centrum, rozhodnutí, jak integrovat formálně nespojité části výpovědi atp. Ke studiu diferenciálních aktivací ve zpracovávání syntaktických a asyntaktických stimulů je tedy potřeba **globální**, ne **lokální** narušení syntaxe (Levitin a Menon 2003, s. 2143). Takového globálního narušení lze dosáhnout tím, že původní nahrávky rozdělíme na krátké segmenty (Abrams et al. zvolili délku okolo 350 ms), které poté promícháme. Tím se v jednotlivých úsecích zachová původní amplitudová obálka i intenzita, ovšem časová oblast, čímž je míněno lineární syntaktické uspořádání jednotek, ale pro nás možná důležitější i rytmus, je změněna k nepoznání. Abrams et al. navíc ve svém experimentu věnovali velkou péči tomu, aby zpřeházené stimuly nebyly samy o sobě nějak akusticky nápadné: pracovali s pseudonáhodným promícháním, při němž se snažili zachovat distribuci přechodů mezi segmenty různých předem definovaných amplitudových úrovní (aby se pozměněná nahrávka nápadně nelišila počtem intenzitních skoků), jakož i distribuci přechodů mezi segmenty různé délky. Též při promíchání důsledně rozdělovali veškeré původně sousedící segmenty, aby promíchaná nahrávka neobsahovala delší koherentní pasáže, které by mohly působit rušivě (Abrams et al. 2011, s. 1509).

Zvolenou zobrazovací metodou byla funkční magnetická rezonance (fMRI), která jako indikátoru pro detekci aktivace neuronů využívá změn v okysličení krve (tzv. *blood oxygen level-dependent activity*, zkratkou *BOLD*). Podobně jako je u digitálních fotoaparátů nejmenší a dále nedělitelnou jednotkou informace o snímaném obrazu pixel, tak je u fMRI prostorové rozlišení určováno tzv. voxely; ve zmiňované studii byly navíc voxely pro účely analýzy poměrně složitou metodou převedeny na tzv. *regions of interest* (ROI, volně přeloženo oblasti zájmu), které nejsou jen rigidní pravidelnou mřížkou a jsou již anatomicky interpretovatelné (Abrams et al. 2011, s. 1510). Statistické zpracování získaných dat probíhalo pomocí metod *pattern analysis*¹⁶, a to jak jednorozměrných, tak vícerozměrných

¹⁶ Anglické slovo *pattern* nemá v tomto kontextu v češtině úplně přiléhavý ekvivalent, nejbližší je snad „vzorec“ ve smyslu opakujícího se uspořádání prvků. Proto také hovoříme o aktivačních vzorcích neuronů

(*multi-variate pattern analysis*, MPA). Každá z nich měla přitom jiné určení: jednorozměrná analýza identifikuje oblasti mozku, které reagují na manipulaci s řečovým či hudebním signálem, kdežto vícerozměrné metody slouží k rozpoznání detailnějších prostorových aktivizačních vzorců v rámci těchto oblastí (Abrams et al. 2011, s. 1507–8).

U řečových i hudebních signálů, původních i promíchaných, se potvrdila již dříve známá výrazná aktivace ve frontálním laloku, konkrétně v Brodmannových oblastech 44 a 45 (BA 44 a BA 45, v rámci levé hemisféry též známé jako **Brocovo centrum řeči**; viz obr. 4) a BA 47 (viz též Levitin a Menon (2003)), a v přední části temporálního laloku (Abrams et al. 2011, s. 1511).¹⁷ Tyto aktivace přitom probíhaly bilaterálně navzdory tomu, že tradičně se o zpracování řeči uvažuje jako o lateralizované funkci soustředěné spíše do levé hemisféry. Proti tomuto lokalizacionistickému pojetí staví Abrams et al. hypotézu, že spíš než o nějaká centra či moduly jde o propojenou distribuovanou síť, která využívá jmenovaných kortikálních a subkortikálních struktur (2011, s. 1515). Jednorozměrná *pattern analysis* pak odhalila, že temporální manipulace se stimuly v obou doménách jsou v těchto strukturách spojeny se změnami o stejné velikosti v síle signálu fMRI (Abrams et al. 2011, s. 1511). Výsledky MPA ovšem nedovolují mezi neurální zpracování syntaxe v řeči a hudbě umístit zjednodušující rovnítko: analýza ukazuje, že detailní prostorové aktivizační vzorce v některých implikovaných oblastech se u řeči a hudby liší do té míry, že lze na jejich základě s úspěšností 70 až 85 % klasifikovat doménu stimulu, jehož temporální manipulace je jejich zdrojem. Jde primárně o BA 44 vlevo, BA 45 vpravo a některá místa v temporálních lalocích (tj. oblasti situované v kůře), v menší míře pak též auditorní mozkový kmen. U něho přitom není jasné, zda jde o přímou reakci na odlišný charakter stimulu, či zda se spíš nejedná o zpětné přizpůsobení těchto nižších funkcí na základě pokynu od korových struktur ve chvíli, kdy vyhodnotí typ vnímaného signálu (Abrams et al. 2011, s. 1515).

Jak poznamenávají sami autoři (Abrams et al. 2011, s. 1516), rozsegmentování nahrávky a následné promíchání ovlivňuje nejen syntax, ale i rytmus, který v percepci řeči i hudby hraje důležitou roli, neboť umožňuje predikovat, v které chvíli bude realizována další relevantní jednotka, a ukotvuje tak oscilaci pozornosti při vnímání a dekodování akustického signálu. To je do jisté míry povzbudivé i pro náš experiment, neboť výše zmíněné změny v síle signálu fMRI tím pádem ukazují také na to, že mozek reaguje podobným způsobem, když se rytmus roztříští v řeči i hudbě. Podpůrným argumentem tu mohou být výzkumy, které dokládají aktivaci řečových oblastí mozku, když subjekt produkuje rytmus, který tvoří součást polyrytmického napětí – ať už je úkolem subjektu držet základní rytmus a polyrytmus vzniká přidáním nahrané sekvence (Vuust et al. 2006),

v mozku.

¹⁷K aktivacím došlo i v některých nekorových strukturách, jako jsou hrbol mezimozkový, mozkový kmen a mozeček.

nebo základní rytmus drží nahrávka a polyrytmus subjekty samy vytvářejí (Vuust et al. 2011). Na druhou stranu jsme ovšem viděli, že detailní aktivační vzorce se mezi oběma doménami částečně liší, tudíž otázka, zda je zpracovávání rytmu v řeči a hudbě propojeno v takovém rozsahu, že spolu obě schopnosti korelují, zůstává otevřena. Dále je potřeba ještě prozkoumat vliv časového rozlišení (tj. délky segmentů) při promíchávání stimulů, jakož i sémantického aspektu řečových stimulů. Některé struktury v mozku totiž reagují na změny srozumitelnosti řeči, a je tedy možné, že výsledné aktivace byly ovlivněny nesrozumitelností zpřeházených stimulů (Abrams et al. 2011, s. 1516).

Je dobré si uvědomit, že sdílené neurální mechanismy pro zpracování řeči a hudby je možné interpretovat nejen jako propojení pouze těchto dvou domén, ale šířeji i jako důkaz toho, že jednotlivé oblasti mozku nejsou tak úzce specializované, jak se dříve soudilo, a zapojují se při řešení širokého pole spřízněných kognitivních problémů. V tomto duchu vysvětlují svá data Koelsch et al. (2002), kteří pomocí fMRI zkoumali percepci hudebních úryvků s příležitostnými nečekanými vsuvkami. Zjistili při tom, že ony vsuvky aktivují celou síť mozkových oblastí (Brocovo i Wernickeovo centrum řeči, temporální oblasti aj.), z nichž jednotlivé sice již byly dříve spojovány s percepcí hudby, ale takto komplexní aktivace se měla za typickou pro zpracovávání řeči (Koelsch et al. 2002, s. 956). Ovšem místo aby pouze upravili škatulky a oproti zbytku kognice nově vyčlenili řeč společně s hudbou, vyvozují z toho otevřenější, méně kategorický závěr, že funkce zmíněné sítě jsou jednoduše méně doménově specifické („domain-specific“), než se dříve předpokládalo (Koelsch et al. 2002, s. 961).

Jaké existují důkazy pro sdílené či naopak oddělené neurální zpracování melodie? Patel (2008) ve svém shrnutí této problematiky vychází z výzkumů, které nejsou založeny na zobrazovacích metodách, ale na práci s jedinci s **vrozenou** a **získanou amúzií**, tj. poruchou percepce či produkce hudby. Získaná amúzie je navozena některými mozkovými lézemi a doložené výskyty jsou poměrně skoupé (Patel 2008, s. 226), takže se těžko shánějí větší vzorky populace s touto vadou a výsledné studie mají spíše kazuistický charakter. Naproti tomu vrozená amúzie je genetickou či vývojovou poruchou, o níž se podle dosavadního výzkumu soudí, že její fyziologickou příčinou by mohly být anomálie v pravém sekundárním auditorním kortexu, jenž mapuje tonální prostor, nebo v pravém inferiorním frontálním kortexu, jenž nejspíše souvisí s krátkodobou pamětí na melodické vzorce. Tyto nedostatky pak pravděpodobně dále ztěžují akvizici hudebních kategorií jako jsou výškové intervaly, čímž se percepční znevýhodnění postižených subjektů ještě prohlubuje, neboť si budují zkreslenou, resp. neúplnou představu o „fonologickém“ systému hudební tradice, která je obklopuje (Patel 2008, s. 235–8). Jedinců s vrozenou amúzií jsou v populaci asi 4 %, jednodušeji se tedy hledají a získávají ke spolupráci a do budoucna představují pro výzkum velmi příhodnou skupinu (Patel 2008, s. 230).

Patel zmiňuje studii provedenou na dvou subjektech se získanou amúzií, v níž šlo o

rozpoznání vět, které se lišily intonačním vyjádřením modality (oznamovací × tázací) či důrazu, a zároveň o rozpoznání tónových sekvencí, který byly syntetizovány na základě stylizace průběhu F0 v původních větách. První subjekt, u něž se amúzie projevovala neschopností zapamatovat si běžné, často slýchané melodie, zvládal obě úlohy dobře, zatímco druhý, jemuž činila potíže identifikace rytmických a melodických vzorců, měl v obou problémy. To podporuje představu, že zpracovávání melodie v řeči i hudbě sdílí stejné neurální prostředky, neboť je narušené buď obojí, nebo ani jedno (Patel 2008, s. 226–8).

Ovšem obdobné výzkumy provedené na větších vzorcích populace s amúzií vrozenou tento závěr na první pohled spíš vyvracejí. Vybraní jedinci si v rozpoznávání intonace vedou většinou velmi dobře, ale citelně selhávají při identifikaci tónových sekvencí. Výsledná úspěšnost je stejně špatná, ať je použita stylizace pomocí stacionárních tónů jako v předchozím experimentu či „klouzavá“ stylizace, která věrněji kopíruje intonační konturu, a to navzdory tomu, že práh pro detekci změny výšky tónu je u vrozených amúziků mnohem nižší, pokud se jedná o glissando a ne skokovou změnu (Patel 2008, s. 231–2).¹⁸ Z této perspektivy se tedy obě činnosti jeví jako neurálně disociované. Patel ovšem navrhuje hypotézu, která onu disociaci převádí do povrchové, behaviorální roviny a zachovává tak možnost neurálního propojení. Nazývá ji hypotézou o hluchotě k melodickým konturám (*Melodic Contour Deafness Hypothesis*); méně krypticky ji můžeme charakterizovat tak, že vrození amúzici jsou sice poměrně spolehlivě schopni určit místo melodické změny, méně už ale její směr (stoupnutí či klesnutí). K tomu Patel dodává druhý předpoklad, a sice že řeč je vůči těmto percepčním nepřesnostem odolná (robustní), kdežto hudba nikoli. Informace obsažené v intonaci jsou totiž často do jisté míry redundantní, tj. odvoditelné z dalších aspektů promluvy či kontextu. Navíc mnohdy stačí určit jen umístění melodické změny – např. u důrazu, na rozdíl od melodémů kódujících větnou modalitu – a u některých osob s vrozenou amúzií se skutečně ukazuje, že s prvním typem užití intonace problém nemají, zatímco s druhým ano (poznají důraz, nepoznají otázku, Patel 2008, s. 233–4). Podobně vyznívají i některé výzkumy na mluvčích tónových jazyků s tímto postižením, u nichž vede k jistým jazykovým deficitům (Jiang et al. 2010). Úhrnem lze říci, že neurální propojení percepce melodie v řeči a hudbě není vyloučené, ovšem na behaviorální (povrchové) úrovni (tedy té, kterou budeme sledovat v experimentální části této práce i my) se nemusí projevit.

Bádání o neurálním propojení řeči a hudby se neomezuje jen na oblast syntaxe (pořadí rytmu, který v popsané studii Abramse et al. hrál roli též) a melodie: existují výzkumy i v dalších oblastech. Již v oddíle 1.2 jsme zmínili práci Stefana Koelsche, jenž se zabývá neurálními korelátami zpracování významu v řeči a hudbě, což mezi odborníky

¹⁸Přesné hodnoty jsou 0,21 a 0,56 půltónu při požadavku na 75% úspěšnost rozpoznání.

nezřídka budí kontroverzi, neboť přisuzování sémantiky¹⁹ hudbě je stále sporné. Jeho poznatky přitom nejsou nezajímavé, s kolegy zkoumal např. to, zda poslech hudby může ovlivnit naše lexikální predikce;²⁰ měřeno metodou evokovaných potenciálů se zdá že ano, vhodně zvolený úryvek má podobný efekt jako sémanticky spřízněná věta (Koelsch et al. 2004, s. 305).²¹ S tím částečně souvisí i téma emocionality, která je na jednu stranu oběma doménám vlastní a tudíž méně kontroverzní než sémantika, ale na stranu druhou snad ještě hůře uchopitelná. Kupříkladu Forde Thompson et al. (2012) se snaží napojit vnímání emocionální²² prozodie na hudební dovednosti a shromažďují mnoho podnětných dřívějších poznatků k tomuto tématu. Jejich vlastní výzkumná metoda je ovšem co se týče reprezentace emocí bohužel poněkud naivní: používají umělé nahrávky „sémanticky neutrálních“ (už to je sporné) vět, na něž se mluvčí snažili vědomě naroubovat různé emoce (Forde Thompson et al. 2012, s. 19031). Tyto a další podobné výzkumy tedy dále rozvíjejí pohled na hudbu a řeč v lidském mozku, v tomto přehledu jim ovšem nebudeme věnovat bližší pozornost.

2.2.2 Kvantifikace strukturních podobností mezi řečí a hudbou

Jak jsme viděli v předchozím oddíle, Abrams et al. (2011) věnovali velkou péči tomu, aby stimuly použité v jejich studii byly navzájem srovnatelné, a to jak na ose hudební × řečové (párování podle emocionality atp.), tak na ose syntakticky validní × narušené (tvorba narušených protějšků pomocí segmentace a důmyslného promíchání původních nahrávek). I v našem experimentu je potřeba dosáhnout jisté úrovně srovnatelnosti stimulů z obou domén, ovšem precizní vyvažování na základě emocionality a dalších faktorů se jeví jako zbytečné, neboť zvolená metodologie percepčních testů koneckonců pracuje s vědomými soudy subjektů a není tedy tak citlivá na variabilitu informačního šumu zpracovávaného mozkiem v pozadí jako zobrazovací metody typu fMRI. Naopak globální narušení syntaxe či rytmu v podobě časového promíchání původní nahrávky sice splňuje požadavek na to, aby se řečové i hudební stimuly, u nichž budou posluchači hodnotit podobnost/identitu, od sebe lišily ve srovnatelných ohledech, ale je pro naše účely příliš hrubé – nelze si ho nevšimnout, tudíž ztrácí smysl zkoumat hypotetickou korelaci mezi citlivostí na změny v obou oblastech. Bylo by tedy dobré najít způsob, jak usouvztažnit jemné melodické a rytmické variace v řeči a hudbě, jejichž percepce bude v centru našeho zájmu.

¹⁹Lze říci, že denotativní/doslovné významy v hudbě nemají místo a je otázka, nakolik jsou její konotativní/asociativní významy formalizované, konvencionalizované a propojené s konkrétními jazykovými jednotkami.

²⁰Příklad: když uslyším heroickou pasáž z Beethovena a poté uvidím slovo „hrdina“ nebo „blecha“, bude můj mozek v obou případech reagovat stejně (Koelsch et al. 2004, s. 302)?

²¹Teoretické rozpracování tématu viz Koelsch (2011).

²²„Emocionální“ je jejich termín, lepší by bylo ovšem užít širšího a v tomto kontextu vhodnějšího pojmu „afektivní“.

Už při pohledu na lingvisticko-fonetickou literaturu týkající se rytmu je jasné, že to není snadný úkol. Intuitivní soudy stran řečového rytmu, částečně založené právě na paralelách s hudbou, se totiž neosvědčily. Nicméně vyvolaly velký zájem a dlouhá desetiletí motivovaly výzkumnou aktivitu, která se pokoušela je empiricky nejprve doložit, poté vyvrátit a nakonec nahradit komplexnějším přístupem, který by fungoval jak v rovině intuice, tak v rovině exaktních měření. Ony intuitivní soudy byly poprvé formulovány Kennethem Pike (Pike 1945) a dále rozvedeny Davidem Abercrombiem (Abercrombie 1965). Vycházejí z představy, že rytmus v řeči je projevem izochronie, tj. (přibližně) uniformních časových intervalů mezi jednotlivými prozodickými jednotkami, přičemž různé jazyky se liší v tom, která prozodická jednotka v nich izochronii podléhá. Podle této teorie mohou být jazyky slabičně izochronní (trvání slabik v rámci promluvy je srovnatelné) nebo taktově izochronní (trvání úseků od přízvuku k přízvuku je srovnatelné); za typického představitele prvního typu bývá uváděna francouzština, u druhého pak angličtina. Někteří autoři též uvažovali o třetím typu izochronie, a to izochronii morové, přičemž mora je sub-slabičná jednotka, která hraje roli např. v prozodii japonštiny.²³

Exaktní měření ovšem výše uvedené druhy izochronie nedokládají (viz např. Roach 1982). Jak shrnuje Patel (2008, s. 176), rytmus v řeči velmi pravděpodobně nesouvisí s žádným druhem periodicity; srov. též Patel et al. (2006, s. 3035): „jeden typ rytmu sice spočívá v izochronii, ale absence izochronie není absencí rytmu.“²⁴ Nepotvrdily se ani hypotézy založené na tom, že při percepci dochází k nelineární transformaci reálných časových intervalů, případně k větší toleranci ohledně přesnosti, z čehož by plynulo, že posluchači izochronii vnímají, aniž by reálně byla přítomná v signálu (viz již výše citovaný výzkum Ily Lehiste, Patel 2008, s. 143–4). Byly proto vyvinuty nové míry, které se snaží intuitivní soudy ohledně rytmické povahy jazyků (jako např. citelný rozdíl v prozodii mezi angličtinou a francouzštinou) kvantitativně uchopit. Mezi ně patří třeba měření podílu vokálních intervalů na trvání promluvy (%V) či směrodatné odchylky trvání konsonantických intervalů (ΔC , viz Ramus et al. 1999), nověji také na variačním koeficientu posledního jmenovaného ukazatele (Dellwo a Wagner 2003; Dellwo 2006). Na základě těchto indikátorů se již daří objektivně oddělit jazyky se subjektivně odlišnou prozodií, jejich nevýhodou ovšem je, že jsou silně vázány na řeč, neboť se opírají o přítomnost hlásek a rozlišení mezi konsonanty a vokály. Pro ty nelze nalézt v hudbě jednoznačné paralely, takže na hudební rytmus tyto míry nelze aplikovat.

Naopak měrou, která se pro charakterizaci řečového rytmu taktéž používá, a přitom je na doméně v podstatě nezávislá, je tzv. normalizovaný index párové variability (*normalized pairwise variability index*, **nPVI**). Obecně řečeno tento index kvantifikuje celkovou mí-

²³Detailní přehled těchto teorií podávají Grabeová a Lowová (2002).

²⁴“[W]hile isochrony defines one kind of rhythm, the absence of isochrony is not the same as the absence of rhythm.”

ru *lokální* variability dané sekvence pozorování na základě vzájemné variability přilehlých párů v této sekvenci.²⁵ Takovou sekvencí pozorování může být např. trvání tónů/vokálů jdoucích v melodii/promluvě za sebou. nPVI je definován následovně (Grabe a Low 2002, s. 520):

$$nPVI = \frac{100}{m-1} \times \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{\frac{d_k + d_{k+1}}{2}} \right|$$

kde:

- m je počet prvků v sekvenci
- d_k je naměřená hodnota sledované vlastnosti (v případě měření rytmu se jedná o trvání, ale mohlo by jít o jinou veličinu) u k -tého prvku sekvence

Charakter nPVI jakožto míry lokální variability si můžeme dobře uvědomit pomocí obr. 5, kde je u dvou hypotetických sekvencí pozorování jedné proměnné porovnán s variačním koeficientem (CV). Ten odpovídá směrodatné odchylce normalizované vůči průměru:

$$CV = \frac{s}{\bar{x}}$$

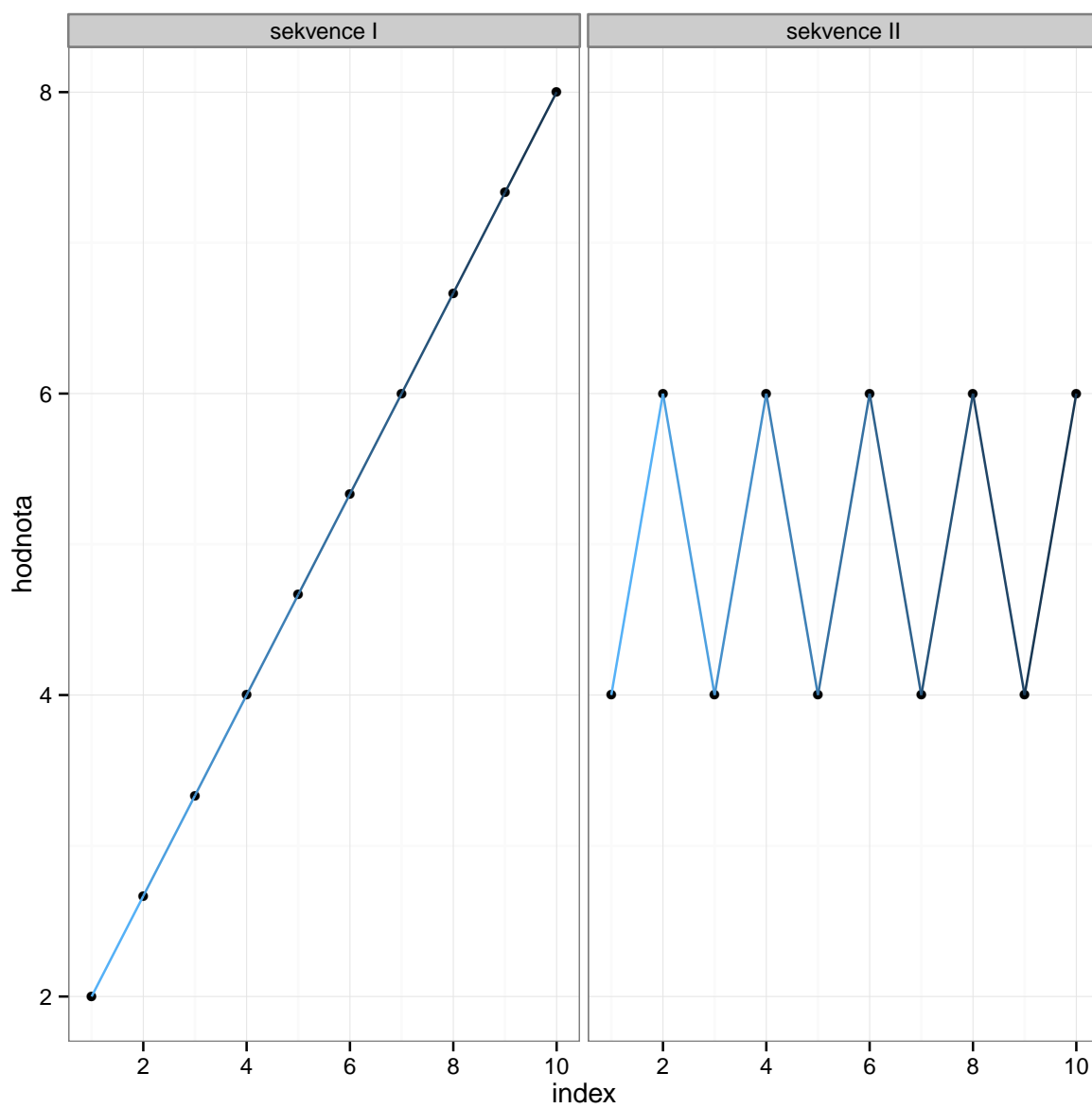
kde:

- s je směrodatná odchylka vzorku
- \bar{x} je aritmetický průměr vzorku

Na okraj poznamenejme, že obě tyto míry jsou relativní, a tudíž lze přímo porovnávat hodnoty, kterých dosahují u různých datových sad. Průměr u obou sekvencí měření v obr. 5 je stejný (5), ale hodnoty u sekvence I jsou mnohem víc rozprostřené, tudíž mají větší směrodatnou odchylku i variační koeficient (40,4 % oproti 21,1 %). U sekvence II jsou naopak hodnoty soustředěné kolem průměru, oscilují kolem něj, ale každý skok mezi dvěma přilehlými měřeními je výraznější než u sekvence I, což se promítá do vyššího nPVI (40,0 % oproti 15,4 %). Porovnání sekvencí z hlediska variability tedy záleží na tom, kterou z obou měř zvolíme. Z toho názorně plyne, že CV i nPVI zachycují jiné druhy variability (globální a úhrnnou lokální) a vztah mezi nimi není přímočarý, neboť u nPVI nezáleží pouze na naměřených hodnotách, ale i jejich pořadí.

Jakožto doménově nezávislou míru použili nPVI pro kvantifikaci rytmu Patel a Daniele (2003) a později v rozšířené studii na podobné téma Patel, Iversen a Rosenberg (2006). Jejich cílem bylo totiž porovnat rytmus v řeči a hudbě, přesněji řečeno zda se rozdílné rytmické vlastnosti angličtiny a francouzštiny odrážejí i v hudbě anglických a francouzských

²⁵Měrou *globální* variability je např. směrodatná odchylka či její normalizovaná podoba (variační koeficient), která nezohledňuje pořadí naměřených hodnot v sekvenci (tj. nepočítá s rozdíly naměřenými mezi přilehlými páry).



Obrázek 5: Dvě hypotetické sekvence měření jedné proměnné a jim odpovídající míry variability globální (variační koeficient, CV) a lokální (normalizovaný index párové variability, nPVI).

skladatelů. Vzorek hudebních úryvků přitom sestavili z prací skladatelů obou národností z přelomu 19. a 20. stol. Toto období bylo totiž podle různých svědectví dobou hudebního nacionalismu, což podle autorů zvyšuje šanci, že se jazyková specifika v hudbě odrazí, neboť skladatelé se snaží hudbě, kterou komponují, vědomě vštípit národní charakter (Patel et al. 2006, s. 3038). Cíleně se soustředili na instrumentální hudbu, aby vyloučili možnost, že jsou melodie šité slově na míru, jak tomu často bývá v lidové hudbě.

Jako materiál pro měření řečového nPVI zvolili trvání sousedních vokálů v korpusu reálných anglických a francouzských promluv; u hudebního nPVI pak trvání sousedních tónů v korpusu melodií, ovšem čistě na základě notace, ne podle reálných nahrávek. Důvodem bylo, že jednotlivé interpretace skladby se mohou v temporální rovině různit a notace byla vyhodnocena jako nejvěrnější zachycení skladatelova záměru (Patel et al. 2006, s. 3038). Výsledky ukazují, že nPVI angličtiny a anglické hudby je statisticky významně vyšší než u francouzštiny a francouzské hudby (55,0 % ku 35,9 % u řečového materiálu a 47,1 % ku 40,2 % u materiálu hudebního).

Na tomto místě je potřeba se na okamžik vrátit k výše popsanému rozdílu mezi globální a lokální variabilitou, tj. mezi CV a nPVI. Vztah mezi nimi totiž sice není přímočarý, ale to neznamená, že žádný neexistuje: část globální variability se pochopitelně musí promítnout na lokální úrovni a záleží jen na řazení hodnot v sekvenci, jak velká část to bude. Bylo tedy potřeba zjistit, zda jsou zjištěné rozdíly v nPVI více dílem celkového rozptylu měření či skokových změn v rámci jednotlivých párů. K odhadu posloužil postup založený na principu metod Monte Carlo: sekvence řečových i hudebních temporálních měření (odvozených z promluv a melodií) byly mnohokrát náhodně promíchány a z těchto permutací byly pro jednotlivé domény i jazyky spočítány nPVI. Povšimněme si, že díky temporálnímu promíchání se výpočet nPVI pokaždé lišil, zatímco globální variabilita byla stejná, neboť jak víme, pořadí měření nemá na hodnotu CV vliv. U každé permutace tedy byl zjištěn rozdíl v nPVI mezi oběma jazyky/hudbami, díky čemuž získali autoři představu o pravděpodobnostním rozdělení možných hodnot lokální variability (nPVI) v závislosti na seřazení sekvencí při stejné hodnotě variability globální (CV). Porovnáním původních zjištěných rozdílů v nPVI (19,1 % u řeči, 6,9 % u hudeb) s těmito rozděleními pak dovedli, že pravděpodobnost dosažení takovýchto a extrémnějších hodnot nPVI na dané hladině CV je 0,01 u řeči, resp. 0,001 u hudeb (Patel et al. 2006, s. 3038). Je tedy vysoce nepravděpodobné, že by zjištěné rozdíly v nPVI nereflektovaly primárně lokální variabilitu zkoumaných sekvencí. Tento poznatek je zajímavý z toho důvodu, že se shoduje s dřívějšími kvantitativně nedoloženými tvrzeními lingvistů, kteří popisují angličtinu jako jazyk, v němž je právě lokální variabilita trvání vokálů nápadná, neboť v něm je velmi výrazná tendence ke střídání plných přízvučných a redukovaných nepřízvučných samohlásek, k čemuž naopak francouzština netíhne (viz odkazy v Patel et al. 2006, s. 3035, 3038).

Patel, Iversen a Rosenberg se ovšem nespokojili s vyhodnocením rozdílů mezi anglickou

a francouzskou prozodií a hudbou čistě na základě rytmu, věnovali se i melodii. Zásadním problémem v tomto ohledu bylo, jak reprezentovat řečovou intonaci tak, aby šla porovnat s hudební melodií, podobně jako tomu učinili u rytmu. Kontury F0 jsou pro tyto účely příliš detailní, zahrnují mnoho variace, kterou percepčně kompenzujeme (např. mikroprozodické jevy) nebo rovnou nevnímáme. Naopak přístup autosegmentální fonologie a z ní odvozeného transkripčního systému ToBI, založený na rozlišení pouze dvou základních tónových úrovních (vysoké H a nízké L), je příliš abstraktní a vrství na empiricky naměřená data interpretaci zatíženou těžko doložitelnou (a zároveň těžko vyvratitelnou) teorií. Autoři se tedy rozhodli zvolit rozumný kompromis mezi oběma extrémy, a tím je pro ně metoda **prozogramu**.

Prozogram vychází z křivky F0 a snaží se ji zjednodušit způsobem, který by odpovídal tomu, jak intonaci vnímá posluchač (Mertens 2004a, 2004b). Klade si tedy podobné cíle jako výzkum percepční stylizace kontur F0 vedený badateli z IPO ('t Hart et al. 1990). Jak popisují Patel et al., podle dosavadního výzkumu na poli percepce intonace se zdá, že při ní dochází ke čtyřem percepčním transformacím:

1. rozdělení kontury F0 do úseků přináležejících jednotlivým slabikám
2. práh pro detekci změny výšky v rámci slabiky
3. pokud je změna výšky detekována, aplikuje se ještě práh pro detekci toho, zda je rychlost změny vnímána jako konstantní či proměnlivá²⁶
4. pokud práh pro detekci výšky změny překročen není, je kontura F0 percepčně integrována na stabilní hodnotu v rámci celé slabiky

(Patel et al. 2006, s. 3036, viz jejich článek pro odkazy na jednotlivé zdrojové studie)

U implementace prozogramu dostupné pro program Praat²⁷ se uživatel postará o první bod (tím, že dodá fonetickou segmentaci hlásek) a algoritmus zajistí zbytek, tj. dosadí pro každou slabiku buď klouzavou konturu přepočítanou podle bodů 2 a 3, nebo stabilní konturu podle bodu 4. Výsledný prozogram obsahuje převážně stabilní úseky, z čehož plyne, že reprezentace intonace v mozku posluchače je překvapivě znatelně podobnější hudební melodii, než by se při pohledu na prvotní křivku F0 zdálo, a jde s ní tedy i jednodušeji kvantitativně porovnávat (Patel et al. 2006, s. 3036). Neméně zajímavý dopad této podobnosti se ovšem týká hypotézy o sdíleném mechanismu pro osvojování zvukových kategorií (SSCLMH), již dříve popsané v poznámce na str. 28. Jde o tzv. statistické učení, při němž jedinec vyvozuje závěry o statistických vlastnostech (rozdělení) opakujících se jevů ve svém okolí, aniž by k tomu využíval explicitní zpětné vazby. Intonace v

²⁶Body 2 a 3 je možné si přiblížit analogií s derivacemi: dosažení prahu v bodu 2 by se určovalo podle průběhu první derivace kontury F0 v rámci slabiky (rychlost změny), v bodu 3 pak podle druhé derivace (zrychlení změny).

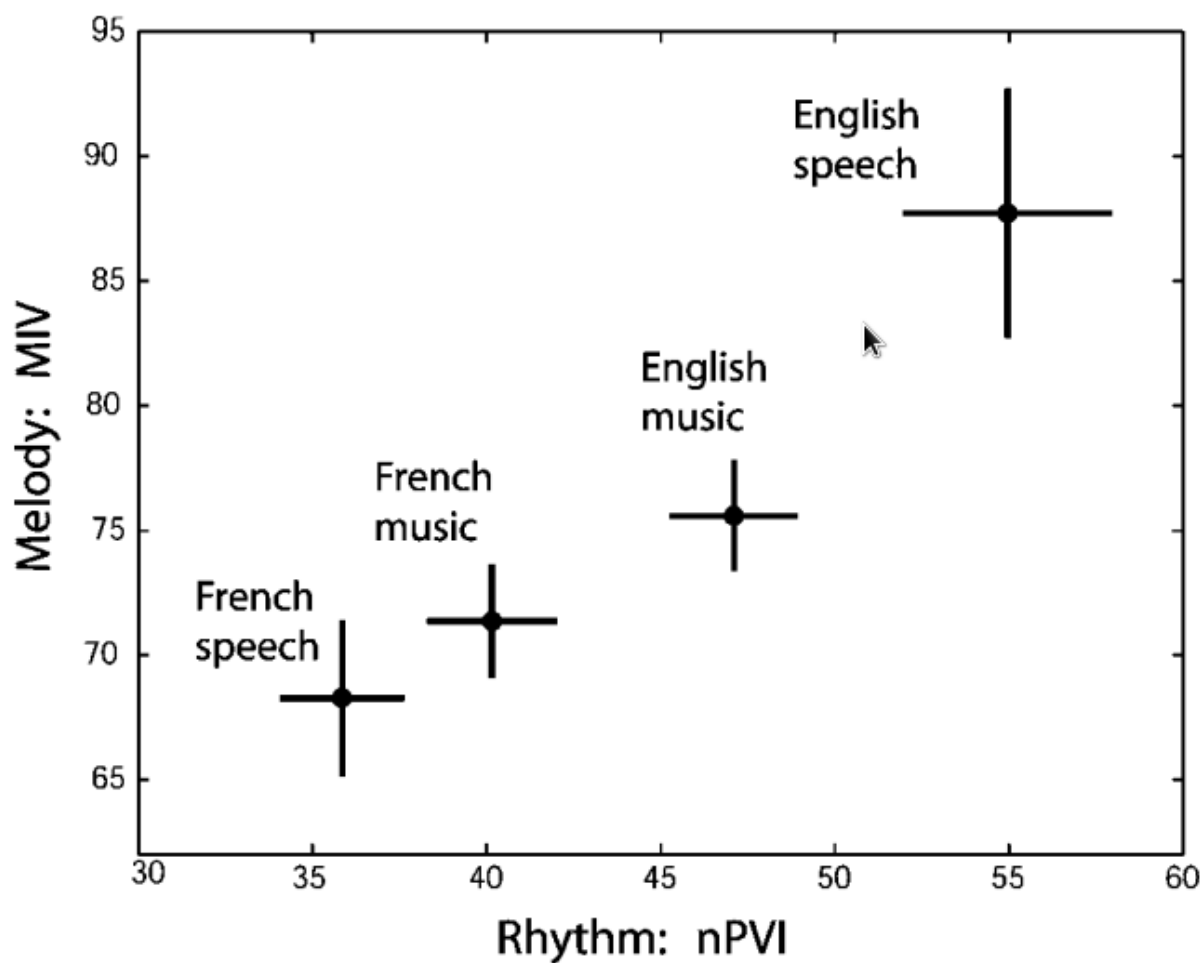
²⁷Volně ke stažení na adrese <http://bach.arts.kuleuven.be/pmertens/prosogram/>.

jazycích na rozdíl od hudby pochopitelně běžně nevyužívá stabilních výškových úrovní a intervalů, byť některé výzkumy naznačují, že jednotliví mluvčí tónových jazyků (např. mandarínštiny) jsou ve volbě intonačních frekvencí při opakovaném čtení seznamu slov pozoruhodně konzistentní²⁸ (Patel 2008, s. 46–7, mluvčí byli opakovaně nahráni ve dvou různých dnech). Při akvizici jazyka si tedy posluchač neosvojuje diskrétní rozdělení přesně vymezených frekvenčních intervalů, nýbrž spíš spojitě rozdělení, z něhož vyvozuje představu o průměru a variabilitě frekvenčních skoků typických pro danou řeč. Onu představu pak promítá do svých očekávání při percepci a do vlastní produkce. Děti jsou přitom k tomuto statistickému učení velmi dobře disponované a je dokázáno, že jsou citlivé na rozdíly v řečových (prozodických) i neřečových tónových sekvencích (Patel et al. 2006, s. 3043). Výše nastíněný mechanismus by tedy mohl představovat i způsob, kterým si skladatelé osvojují představu o rozpětí tónových intervalů typických pro jejich jazyk, kterou pak mohou později aplikovat při komponování (Patel 2005, s. 62).

Díličí výsledky studie Patela, Iversena a Rosenberga ukazují, že tomu tak skutečně nejspíš je: ačkoli průměrná velikost intervalu mezi dvěma sousedícími stabilními tóny (viz bod 4 výše) v prozogramu je v angličtině i francouzštině stejná, variabilita jednotlivých intervalů (měřená variačním koeficientem) je v angličtině statisticky významně vyšší, a podobně tomu je i u obou hudebních vzorků (Patel et al. 2006, s. 3042). Zjištěné rytmické a melodické charakteristiky obou řečí a jazyků lze pak promítnout do dvourozměrného prostoru (rytmus \times melodie), v němž je možné všechny rozdíly intuitivně uchopit a vizuálně nahlédnout (viz obr. 6). Jak poznamenávají sami autoři, není překvapivé, že právě obě hudby jsou si v tomto prostoru blíže než jazyky, neboť hudba je ohniskem uměleckého výrazu a je tudíž více předmětem vědomé manipulace a kreativní práce s kulturními normami, nejen domácími, ale i zahraničními – inspiraci se meze nekladou. Při vší této proměnlivosti, volnosti a různorodosti je pozoruhodné, že posun od francouzské k anglické hudbě nastává ve zvoleném rytmicko-melodickém prostoru podobným směrem jako posun od francouzštiny k angličtině (Patel et al. 2006, s. 3041).

Metoda prozogramu a vzorec nPVI tedy umožňují uchopit a kvantifikovat rytmickou a melodickou variabilitu, a to srovnatelným způsobem u řeči i hudby. Zdůrazněme zde ještě jednou, že nPVI se z lingvistického hlediska shoduje s intuitivními soudy o rozdílech v prozodii mezi různými jazyky, a není tedy jen slepou kvantifikací některých temporálních aspektů řeči, nýbrž minimálně koreluje s jednou či více dimenzemi řečového rytmu, které jsou percepčně relevantní. Odráží tak nepřímou lidskou percepci rytmu. Podobně prozogram jakožto zjednodušení intonační křivky je motivovaný percepčně.

²⁸Další studie ovšem zase naznačují, že mluvčí netónových jazyků (konkrétně angličtiny) jsou v tomto ohledu sice méně konzistentní, ale ne nijak výrazně: 3/4 půltónu u mluvčích mandarínštiny a vietnamštiny proti 1 půltónu u mluvčích angličtiny (Patel 2008, s. 47). Celá problematika si tudíž žádá dalšího zkoumání.



Obrázek 6: Prostor indukovaný rytmickými (nPVI trvání sousedních vokálů/tónů) a melodickými (variabilita melodických intervalů) ukazateli řeči a hudby, spolu s polohou angličtiny, francouzštiny a anglické a francouzské hudby v něm. Úsečky naznačují směrodatné chyby. Převzato z (Patel et al. 2006, s. 3041).

2.2.3 Právě postřehnutelné temporální a frekvenční rozdíly (JND)

Neméně důležitým aspektem celé problematiky je ovšem i otázka, jak malé temporální a frekvenční rozdíly jsou ještě percepčně relevantní. Jinými slovy, jaký je **právě postřehnutelný rozdíl** (*just noticeable difference*, **JND**) v temporální a frekvenční doméně? Jak uvádí Nootboom (Nootboom 1999, s. 9), práh temporálního rozlišení dvou *izolovaných* řečových či neřečových zvuků o trvání od 40 do 250 ms je od 5 do 15 %; u delších a kratších zvuků se přesnost odhadu snižuje a zvuky kratší než 20 ms ze subjektivního hlediska zdá se trvání nemají. U délkového kontrastu sousedících vokálů v plynulé řeči (tj. ne izolovaných zvuků), které je základem pro výpočet nPVI, je ale odhad trochu složitější, je třeba si uvědomit, že „sousední“ vokály nemusejí následovat bezprostředně po sobě, bývá mezi ně vmezen jeden či více konsonantů, které odhad ztěžují. Nootboom poznamenává, že JND trvání hlásek ve spojitě řeči se z praktických důvodů zkoumá špatně, a referuje o výzkumném paradigmatu, podle něž se na tento práh usuzuje na základě fonémického rozlišení krátké a dlouhé samohlásky ve větě (ve dvojici slov, která z hlediska vokálního délkou tvoří minimální pár). V jím zmiňovaném experimentu byl přechod od subjektivní percepce krátkého k dlouhému vokálu v závislosti na jeho objektivní délce modelován jako kumulativní distribuční funkce s průměrem 90 ms a směrodatnou odchylkou okolo 5 ms, z textu už bohužel ale není jasné, zda byla za hodnotu JND považována právě tato směrodatná odchylka či byl proveden nějaký další výpočet. Samotná směrodatná odchylka jako JND totiž nedává moc smysl, lepší by bylo vzít dvě směrodatné odchylky na oba směry od průměru, neboť takový interval pokryje přes 95 % přechodu mezi úrovněmi, v nichž jsou si posluchači délkou jisti, tj. mezi nimiž spolehlivě rozlišují a jejich soudy v nich nejsou dílem náhody. Dohromady by tedy práh odpovídal čtyřem směrodatným odchylkám, tj. přibližně 20 ms, ale navíc je potřeba počítat s tím, že pokud mají obě délky v jazyce posluchačů fonologicky distinktivní status, dochází k deformaci percepčního prostoru vlivem kategoriální percepce a výsledky tak mohou být navíc zkreslené. Nootboom nicméně zmiňuje i jiná výzkumná paradigmatata, v nichž posluchači plnili náročnější úkoly, které nebyly přímo závislé na fonologii daného jazyka; v nich došli badatelé k hodnotám 25 ms a více, někdy dokonce až 40 ms, přičemž tyto odhady by měly být konzervativnější. Pols (1999, s. 11) naopak překvapivě uvádí hodnotu výrazně nižší, a to 5 ms ze základu 50ms, přičemž JND se v závislosti na trvání tohoto referenčního zvuku podle něj mění.

Co se citlivosti ve frekvenční oblasti týče, 't Hart, Collier a Cohen ve svém přehledu uvádějí, že v rozsahu typickém pro anglický mužský hlas se JND výšky hlasu u syntetizovaných vokálů pohybuje mezi 0,3 až 0,5 %, ²⁹ což je srovnatelné s citlivostí na změny výšky sinusoidálních zvukových vln. Dodávají ale, že někteří autoři přišli s mnohem vyššími prahy, např. 5 % kolem 150 Hz s resyntetizovanou řečí či 4 % kolem 195 Hz s přirozeným

²⁹Pols (1999, s. 10–11) zmiňuje stejné hodnoty s tím, že při zhoršených poslechových podmínkách či kvalitě signálu se JND rapidně zvyšuje.

hlasem ('t Hart et al. 1990, s. 28).

Z velké variability údajů v tomto přehledu plyne, že se na ně nelze slepě spoléhat při odhadu, jaké temporální či frekvenční rozdíly budou na hranici rozlišitelnosti a jaké naopak již hodně nápadné. Je patrné, že výsledná zjištění jsou silně závislá na konkrétním pojetí experimentálního úkolu, kontextu stimulů i stimulech samotných. Ve výzkumu, který se pohybuje na hranici difference limen v těchto oblastech, tedy nic nenahradí důkladné pilotní testování přímo se stimuly, které v něm budou použity.

Bylo by zajímavé spočítat, jak jsou na tom z hlediska JND rytmické a melodické rozdíly pozorované Patelem, Iversenem a Rosenbergem. Jak je vidno již z obr. 6, nejmenší mezikulturní rozdíl je mezi francouzskou a anglickou hudbou. Je i percepčně relevantní? A co rozdíl mezi jazyky? K tomu by ovšem byl nutný přístup k původním datům, neboť publikované míry nPVI a CV jsou již měrami relativními, kdežto zde potřebujeme nejprve vyjít z hodnot absolutních. Kromě toho JND pro porovnání rozdílu ve velikosti dvou různých změn (oproti tradiční výše popsané JND, kde jde jen o postřehnutí změny) budou pravděpodobně zase jiné hodnoty, navíc znovu silně závislé na konkrétním charakteru stimulů. Lépe než teoretický výpočet by tedy bylo provést se srovnatelným materiálem rovnou percepční experiment. Jeden takový popisuje Hannonová (2009) a podíváme se na něj blíže v následujícím oddílu.

2.2.4 Řeč a hudba v percepci

Viděli jsme, jak se vztah mezi řečí a hudbou dá zkoumat z pohledu neurologického i z hlediska objektivní kvantifikace akustických charakteristik. Podívejme se nyní na závěr teoretické části naší práce na některé studie, které problematiku zkoumají z perspektivy dovedností mluvčího/posluchače, tj. jeho percepce a produkce, a jsou tak našemu experimentu metodologicky na první pohled nejbližší.

Nejprve zmiňme experiment Hannonové (2009), který je percepčním protějškem snah Patela et al. (2006) o kvantifikaci vztahu mezi rytmem ve dvou jazycích a jejich spřízněných národních hudebních tradicích. Hannonová v něm z jejich předchozích zjištění přímo vychází a snaží se ověřit, zda objektivně zjištěné rozdíly v hudebním rytmu (měřené pomocí nPVI) jsou i percepčně postřehnutelné, tj. zda na jejich základě mohou posluchači uhodnout kulturně-jazykové prostředí, v němž daný hudební vzorek vznikl. Jako stimuly použila instrumentálně provedené melodie z francouzských a anglických lidových písní, u nichž se dá předpokládat blízká spřízněnost s jazykovou prozodií, neboť byly původně určeny pro zpěv. V experimentu 1 vybrala vzorek melodií tak, aby se rytmicky podle nPVI co nejvíce lišily a rozdíl byl tedy co nejnápadnější, v experimentu 2 pak podle toho, aby průměr vzorku co nejlépe reflektoval populaci písní, z níž vybírala. V druhém experimentu si tedy byly anglické a francouzské písně v průměru navzájem rytmicky podobnější. Oba experimenty navíc provedla se dvěma variantami stimulů: ve variantě a) byly melodie v

nezměněné podobě, kdežto ve variantě b) byly realizovány na jediné notě (monotónně), čímž z nich byla zcela vyřazena informace o melodické kontuře a zachovány byly pouze rytmické charakteristiky.

Aby minimalizovala vliv toho, že posluchači mohli některé melodie znát (jsouce Američany, zvláště ty anglické), používala Hannonová v rámci experimentů pro kulturní příslušnost melodií místo slov „angličtina“ a „francouzština“ smyšlených jmen „Navi“ a „Latu“. Zároveň měli posluchači ohodnotit, jak moc jsou jim melodie povědomé, aby bylo možné odhadnout, zda úspěšnost při kulturní klasifikaci melodií neplyne jen z toho, že je subjekty znají. Posluchači vždy nejprve absolvovali cvičnou fázi, v níž u každého stimulu dostali po vlastním klasifikačním odhadu zpětnou vazbu se správnou odpovědí, a následně ostrou fázi.

U obou experimentů s oběma variantami stimulů byla klasifikační úspěšnost statisticky významně nad úrovní náhody. V experimentu 1 (s rytmicky odlišnějšími melodiemi) byla tato úspěšnost vyšší, ale z pohledu variant stimulů (s původní konturou či monotónní) se úspěšnost nijak výrazně nelišila (1a: 77,3 %, 1b: 77,1 % proti 2a: 68,4 %, 2b: 70,5 %). Analýza rozptylu potvrdila hlavní efekt u zvoleného vzorku melodií (experiment 1 \times experiment 2), ne už tak u varianty stimulů (a \times b) a neodhalila žádnou interakci (Hannon 2009, s. 406–7). Oba tyto výsledky podporují hypotézu, že se posluchači skutečně rozhodovali podle rytmických rozdílů: čím víc jsou sady stimulů od sebe rytmicky vzdálené, tím je úspěšnost vyšší, a zároveň se ukazuje, že na úspěšnost nemá vliv, pokud ve stimulech skutečně ponecháme pouze rytmickou informaci a melodii smažeme.

Co se týče povědomosti zvolených hudebních melodií, u anglických písní byla vskutku převážně vyšší, kromě experimentu 2b, kde se trend sice obrátil, ovšem statisticky nevýznamně. S úspěšností klasifikace povědomost signifikantně korelovala pouze v experimentu 1a (rytmicky výrazně odlišné stimuly se zachovanou melodickou informací). Konkrétněji, pro každou ze čtyř variant experimentu byl sestaven lineární model tendence klasifikovat danou píseň jako anglickou (závislá proměnná) s nPVI a povědomostí stimulu jako prediktory (nezávislé proměnné); explanatorní přínos povědomosti v rámci modelu byl statisticky významný pouze v onom prvním případě (1a). Podle všeho tedy posluchači snad opravdu písně klasifikovali na základě jejich rytmických charakteristik a ne podle toho, zda některé z nich předem znali.

Hannonová pro své výsledky nabízí podobné vysvětlení jako Patel et al. (2006), a to že posluchači nevědomky shromažďují statistické informace o kulturně-specifických rytmických vlastnostech jazyka, potažmo hudby ve svém okolí, které pak využívají při klasifikaci rytmických sekvencí. Uvádí, že její výzkum ukazuje, že posluchači jsou v hudbě citliví na rytmické ukazatele typické pro jejich jazyk, čímž nepřímo podporuje hypotézu, že percepce rytmu v obou doménách sdílí stejný kognitivní mechanismus. Také ovšem připouští, že oba mechanismy mohou být oddělené, jen natrénované na rytmicky podobných datech

z hudby a řeči, a tudíž zdánlivě sprážené (Hannon 2009, s. 407). Opatrnost je vsutku namístě, uvědomme si, že Hannonová použila pouze hudební stimuly, místo aby zařadila též stimuly řečové a zkoumala, zda u jednotlivých respondentů lze nalézt korelaci v klasifikačních výkonech v obou doménách, která by spíše ukazovala na sdílený mechanismus či alespoň možnost přenosu percepčních dispozic (vrozených či získaných) mezi oběma doménami. Pro nás je ovšem důležité hlavně to, že variabilita nPVI se zdá být korelátem vnímaných rozdílů v rytmu.

Jako poměrně intuitivní se jeví předpoklad, že jedinci s vyvinutým hudebním sluchem dokáží lépe postřehnout i jemné rozdíly ve zvukové stránce jazyka, a jsou tudíž nadanější pro akvizici fonologie cizího jazyka (L2) v dospělosti. Slevc a Miyake (2006) se rozhodli tuto intuici empiricky otestovat a prozkoumat, zda nemají hudební dovednosti vliv i na osvojování některých nezvukových aspektů L2. Zjistili, že hudební zdatnost predikuje lepší výsledky v akvizici fonologického systému cizího jazyka (jak z hlediska produkce, tak percepce), ale na osvojování syntaxe či lexikální zásoby vliv nemá.

Specifický byl Slevcův výzkum zejména tím, že pro vyhodnocení hudebních dovedností subjektů využil psychometricky validního standardního praktického testu (Wing 1968), na rozdíl od jiných studií, které často spoléhají pouze na subjektivní sebehodnocení respondentů, případně doplněné informací o absolvovaném formálním hudebním vzdělání. Také se pečlivě snažil zohlednit veškeré další faktory, o nichž je známo, že mají na osvojování L2 vliv, a kdyby nebyly zahrnuty do celkového modelu, mohly by efekt hudební proměnné zkreslit. Jedná se zejména o věk příchodu do cizí země a délku pobytu v ní (obojí předpokládá, že je mluvčím imigrantem), pravidelnost, s níž mluvčí cizí jazyk používá, a kapacita fonologické krátkodobé paměti (Slevc a Miyake 2006, s. 676).

Respondenty byli rodilí mluvčí japonštiny, kteří emigrovali do USA. Test fonologické percepce spočíval v identifikaci správného slova z minimálního páru, který se lišil fonémem, jež v angličtině činí Japoncům problémy.³⁰ V druhé části si pak pokusné osoby poslechly nahrávku krátkého vyprávění a paralelně měli v jejím textu u vyznačených slov zaznamenat, zda se v nich vyskytla výslovnostní chyba. Test fonologické produkce spočíval ve snaze naopak vyslovit slova, která jsou členy minimálních párů, a přečíst krátký ucelený anglický text. Výkon respondentů byl hodnocen v první části třemi, ve druhé dvěma rodilými mluvčími angličtiny.

Jak již bylo výše naznačeno, statisticky významná část rozptylu v naměřených datech o fonologické produkci a percepci v cizím jazyce byla vysvětlena zahrnutím proměnné hudební zdatnosti do lineárního regresního modelu, a to i při zohlednění ostatních relevantních proměnných, které mají na akvizici L2 vliv. Pozornosti hodné přitom je, že autoři též pro zajímavost požádali mluvčí o sebehodnocení stran hudební zdatnosti, a korelace

³⁰Slovo bylo umístěno do věty, v níž oba členy minimálního páru byly sémanticky přijatelnými možnostmi.

tohoto faktoru s výkony v cizím jazyce byla o poznání slabší. Oproti míře hudební zdatnosti stanovené na základě psychometrického testu dokonce jeho explanatorní přínos v rámci modelu nebyl ani statisticky významný (Slevc a Miyake 2006, s. 679). Slevc zjištěný efekt vysvětluje tím, že lidé, kteří se nějaký cizí jazyk učí až v dospělosti, se své věkové znevýhodnění snaží kompenzovat tím, že se při akvizici snaží využít různé neřečové dovednosti. Dodává ale, že kauzální vysvětlení korelace mezi dovednostmi v hudbě a L2 může spočívat v nějakém třetí, základnější příčině, kterou by mohly být např. lepší auditorní predispozice a na ně vázaná schopnost detekovat jemné akustické rozdíly (2006, s. 680).

Studie Dankovičové et al. (Dankovičová et al. 2007) na straně hudebních faktorů také kromě formální průpravy zohledňuje i prakticky prokázané dovednosti a zkoumá jejich korelaci s výkony subjektů v oblasti intonační analýzy angličtiny (podle klasifikace zavedené tzv. britskou školou). Výzkum je jedinečný tím, že skupina respondentů je z hlediska cílové zkoumané dovednosti poměrně homogenní: všichni jsou studenty stejného ročníku stejného oboru s lingvistickým zaměřením a intonační analýze se všichni věnovali v rámci fonetického modulu svého studijního programu. Lze tedy předpokládat, že k osvojení těchto schopností měli všichni stejné podmínky, a případné rozdíly ve výsledných výkonech tak budou zapříčiněny jinými faktory, např. právě hudební anamnézou. Na jejím základě, konkrétně na základě dosaženého formálního hudebního vzdělání, byly pokusné osoby rozděleny do dvou skupin, na „hudebníky“ a „nehudebníky“.

Studie popisuje dva experimenty, přičemž druhý se od prvního lišil zejména tím, že obsahoval pilotní fázi, v níž se předem testovala složitost jednotlivých experimentálních úkolů, aby se předešlo stropovým efektům u těch, které se ukázaly jako příliš lehké (Dankovičová et al. 2007, s. 195). Obecně řečeno, v obou experimentech spočívaly hudební úkoly v posuzování směru změny výšky tónu a v testování paměti na melodie (pár melodií se vždy lišil jedním tónem, který bylo potřeba identifikovat a navíc stanovit, zda byl v druhé melodii výše či níže). Intonačně-analytická část pak zahrnovala úlohy diskriminační (stejná × odlišná kontura) a identifikační (umístění intonačního nukleu v rámci věty, typ melodického akcentu v něm).

Detailními výsledky Dankovičové studie se zde nebudeme zabývat, není na to prostor, a odkážeme čtenáře na její původní článek. Velmi podrobně rozebírá ohledy, v nichž se hudebně zdatní jedinci od těch méně zdatných v jednotlivých dovednostech liší, které úkoly jsou naopak pro obě skupiny srovnatelně jednoduché či těžké, a jak s tím vším souvisí hudební vzdělání. Výše nastíněné úkoly jsou navíc rozděleny do podtypů (např. podle délky cílové melodické sekvence či pozice intonačního nukleu ve větě), což umožňuje autorům přesněji identifikovat klíčové úlohy, které přispěly ke stratifikaci úspěšnosti. Globálně můžeme říct, že „hudebníci“ „nehudebníky“ předčili ve většině hudebních i intonačních úloh (Dankovičová et al. 2007, s. 211). Ať už byla hudebnost měřena pomocí deklarované formální průpravy či výkonu v praktických testech, byla zjištěna silná pozitivní korelace

se skórem v intonačně-analytických úlohách (Dankovičová et al. 2007, s. 212). Jak ale Dankovičová správně poznamenává, nejpravděpodobnější scénář je, že lidé s hudebními vlohami (a tedy predispozicemi pro dobrý výkon v praktických testech) v životě tíhli i k získání formálního hudebního vzdělání, a vzdělání tedy není náhradou nadání. Příznačný je v tomto ohledu případ jednoho „nehudebníka“ z druhého experimentu, který měl s hudební praxí rozsáhlé bytí neformální zkušenosti, a jeho výkon byl srovnatelný s „hudebníky“; naopak několik „hudebníků“ dosáhlo výrazně horších výsledků než zbytek skupiny (Dankovičová et al. 2007, s. 211).

Sadakata a Sekiyama (2011) se podobně jako Slevc věnují percepční citlivosti na fonologii cizího jazyka, a to v doméně temporální a spektrální (geminace a kvalita japonských konsonantů, kvalita nizozemských vokálů), ovšem zkoumají již jen korelaci s hudebním vzděláním. Nicméně velmi šikovně tematizují distinkci, s níž jsme se implicitně setkali již u Dankovičové: rozdíl mezi auditorní **diskriminací** a **identifikací**. Podle nich jde o dvě různé úrovně percepce, přičemž první – diskriminace – probíhá víceméně automaticky, ani nevyžaduje vědomou pozornost (tj. na temporální, výškové, spektrální či intenzitní variace v akustickém signálu náš mozek reaguje, aniž bychom se na ně soustředili); naopak druhá – identifikace – vyžaduje přístup k předem vytvořeným kategoriím akustických událostí, jejichž osvojení je netriviální záležitostí (Sadakata a Sekiyama 2011, s. 2). Obecně se ukazuje, že hudební výchova výrazným způsobem formuje právě onu první, preatentivní a diskriminační fázi; v rámci tohoto experimentu pak bylo patrné, že má mnohem větší vliv na percepci (jak diskriminaci, tak identifikaci) temporálních rysů, tedy detekci japonské geminace, než těch spektrálních (Sadakata a Sekiyama 2011, s. 8). Jako možnou příčinu tohoto rozdílu autoři uvádějí fakt, že temporální variabilita se vyskytuje v obou doménách, jazykové i hudební, a je tedy sdíleným rysem, kdežto spektrální charakteristiky vokálů, specifické pro jazykové akustické systémy, nemají v hudbě jasnou paralelu.

Další oblastí, která se ke zkoumání spojitostí mezi percepcí řeči a hudby přímo nabízí, jsou tónové jazyky. Jak jsme uvedli již výše, hlasovou melodii využívají způsobem, který je ze všech jazyků hudbě nejbližší. Zároveň se zdá, že mluvčí tónových jazyků častěji disponují schopností absolutního sluchu, jež úzce souvisí i s hudbou (Patel 2008, s. 47; Deutsch et al. 2006).³¹ Právě tento aspekt – konkrétně hypotézu o relevanci absolutního sluchu pro percepci tónových jazyků – se rozhodli na materiálu z mandarínštiny probádat Lee a Lee (2010). V rámci experimentu měli mandarínsky hovořící hudebníci nejprve identifikovat hudební tóny (tj. jmenovat odpovídající noty) ve třech různých ténbrech

³¹Deutsch et al. naznačují, že v obou případech se jedná o schopnost asociovat s tónem (jazykovým či hudebním) pojmenování a přes toto pojmenování si daný tón nezávisle na kontextu vybavit. Ovšem pozor, ačkoli se stejně jmenují, bylo by unáhlené považovat tóny v jazyce a v hudbě za jeden a ten samý typ entity. Je tedy namístě se ptát, zda ono připodobnění „absolutního sluchu“ v hudbě a řeči příliš nespolečá na prvotní metaforu, kterou se slovo „tón“ zavedlo jako termín pro popis jazyků.

bez referenčního tónu. Ve druhé fázi se pak snažili identifikovat krátké stimuly vytvořené na základě mandarínských tónů, z nichž byla ovšem odstraněna dynamická variace F0 a údaje, o nichž se běžně soudí, že jsou nutné pro normalizaci vnímaného zvuku vůči intonačnímu rozsahu mluvčího. Na rozdíl od kantonštiny, v níž jsou některé tóny odlišeny prostou výškou, liší se mandarínské tóny primárně konturami; nicméně i na jejich percepci může mít vliv výška F0, takže vymazání dynamické (konturové) informace by nemělo identifikaci úplně znemožnit, jen ztížit (Lee a Lee 2010, s. 482). Lee a Lee predikovali, že pokud je absolutní sluch propojen s akvizicí tónového jazyka, měly by být výkony subjektů v obou typech úloh korelovány; žádná korelace ovšem nalezena nebyla (Lee a Lee 2010, s. 488). Z toho by bylo možné vyvodit, že oba mechanismy spolu nijak zvlášť nesouvisí, nicméně na vině by mohla též být nevhodná volba experimentálních úloh, která vedla k příliš malé variabilitě v responzních datech. Zdá se totiž, že první (hudební) úloha byla pro subjekty příliš jednoduchá, což vedlo ke stropovému efektu, zatímco ta druhá, jazyková, byla naopak velmi náročná a vedla k výsledkům, které se sotva přehoupaly přes úroveň náhody (Lee a Lee 2010, s. 489).

Snad lépe a jednodušeji se podobné hypotézy zkoumají na již výše zmíněné kantonštině, která přímo disponuje tóny odlišenými pouze výškou, a není v ní tudíž potřeba provádět složité manipulace směřující k potlačení konturové informace v průběhu F0, které mohou celou identifikační úlohu nepřiměřeně ztížit. O takovém výzkumu informují Mok a Zuo (2012). Testovali identifikaci standardních kantonských tónů (v jednoslabičných slovech a resyntetizovaných čistě tónových průbězích) u tří skupin posluchačů. První tvořili běžní mluvčích kantonštiny, druhou mluvčí kantonštiny, v jejichž řečové produkci některé tóny splývají,³² a třetí cizinci; ve všech těchto třech skupinách byli přítomni jak hudebníci, tak nehudebníci. Výsledky ukázaly, že hudební praxe přispěla k významně vyšší úspěšnosti pouze u cizinců, u rodilých mluvčích neměla žádný vliv, a to ani u těch, kteří běžně ve vlastních projevech některé standardní tóny nerozlišují a jsou tedy k potenciálním záměnám náchylnější.

Mok a Zuo z toho vyvozují, že percepce řeči a hudby jsou na sobě do velké míry nezávislé, ovšem znovu podceňují, resp. zmiňují jen letmo, možnost stropového efektu, který mohl zabránit většímu rozvrstvení úspěšnosti mezi rodilými mluvčími. Úlohy byly totiž poměrně jednoduché, i cizinci v nich dosahovali úspěšnosti nad 90 %. Navíc přímo z podstaty věci plyne, že mluvčí daného jazyka bude na percepci jeho základních zvukových kategorií velmi dobře natrénovaný, ať hudebníkem je či není, jinak by měl potíže se ve své rodné řeči domluvit. Můžeme si to představit tak, že méně hudebně nadanému jedinci může osvojení těchto kontrastů trvat déle, ale je krajně nepravděpodobné, že by se mu

³²Splývání (anglicky tzv. *merger*) některých tónů se v současné době prosazuje jako jazykový trend u části hong-kongské populace mluvčích kantonštiny, nejspíš pod tlakem imigrace z Číny, kontaktem s jinými varietami čínštiny a zejména rostoucím vlivem mandarínštiny (Mok a Zuo 2012, s. 2711).

je nepodařilo časem osvojit se spolehlivostí těžko rozlišitelnou od jedince hudebně nadaného, když je svému rodnému jazyku nepřetržitě vystaven. Koneckonců, jak jsme viděli již výše v diskusi o osobách s vrozenou amúzií, je naopak možné mít dokonce hudební deficit a navzdory tomu nevykazovat žádné nedostatky v oblasti jazykového zpracování prozodické informace, neboť řečový signál je robustní a nabízí mnoho záchytných bodů, bohatý sémantický kontext nevyjímaje. Zkrátka, konvenční řečové úkony jsou něčím, co ovládáme až příliš dobře; aby bylo možné zkoumat vliv hudebních vloh na percepci řeči, bývá přínosné sáhnout k do jisté míry „umělým“ úkolům, jako je např. formální intonační analýza zvolená pro tento účel Dankovičovou. Takovým „umělým“ úkolem byla i identifikace kantonských tónů pro cizince a fakt, že právě jim v úkolu hudební průprava výrazně napomohla, je naopak zajímavý velmi, což Mok a Zuo příliš nezdůrazňují.

I jiné body argumentace autorů jsou sporné – např. to, že ve svém vzorku našli mluvčí kantonštiny, kteří jsou sice hudebníky, ale vykazují tendenci ke splývání tónů (Mok a Zuo 2012, s. 2717, 2718), nutně na separaci řečové a hudební percepce neukazuje. Nelze totiž automaticky předpokládat, že ke splývání tónů dochází kvůli deficiencím v percepci řeči. Jak doložil ve své průkopnické sociolingvistické studii již Labov (1963), důvody, pro něž skupiny mluvčích volí konkrétní variantu v rámci jedné sociolingvistické proměnné, jsou sice poměrně různorodé, ovšem většinou souvisí se sociální identitou jedince spíše než s nedostatky ve sluchové percepci. Jinými slovy, kantonští hudebníci se splývajícími tóny nejsou nutně lidé, kteří v hudbě slyší „dobře“ a v řeči „špatně“, ač se to Mok a Zuo snaží naznačit a doložit tak, že jsou obě domény oddělené.

Na závěr tohoto oddílu zmiňme alespoň zběžně metodologicky velmi zajímavý experiment Fedorenkové et al. (2009). Autorka se rozhodla explicitně testovat Patelovu **SSIRH** (hypotézu o sdílených prostředcích pro syntaktickou integraci) prostřednictvím percepčního experimentu se zpívanými větami, tj. stimuly, které na rozdíl od ostatních výše zmíněných experimentů vyžadují simultánní percepci řeči i hudby. Manipulovala u nich se dvěma proměnnými, které měly ovlivnit jejich syntaktickou komplexitu, a tudíž i kognitivní náročnost jejich zpracování. Na straně jazyka to byla proximita na sobě závislých větných členů (čím jsou závislé členy blíže k sobě, tím je věta kognitivně jednodušší), na straně hudby pak zachování tóniny na kritickém místě melodie oproti vybočení z ní. Kromě toho zařadila i stimuly, kde bylo místo výšky manipulováno s hlasitostí, aby se ujistila, že posluchači skutečně reagují na syntaktickou komplexitu hudební manipulace (tj. vybočení ve vztahu k předem známému systému) a ne pouze na čistě akustickou povrchní odlišnost. Sledovanou proměnnou pak byla úspěšnost odpovědí na otázky související s propozičním obsahem předložených vět.

Fedorenková předpokládala, že pokud je syntaktická integrace v obou doménách propojená, měly by stimuly, které jsou syntakticky složitější/narušené v obou ohledech, vyústit ve snížení úspěšnosti, které nebude pouhým součtem efektů obou faktorů v izolaci (tzv.

additive factors logic, Fedorenko et al. 2009, s. 3). Jinými slovy, faktory podle ní nejsou nezávislé a mělo by dojít k interakci, protože se stejné mozkové struktury či mechanismy snaží současně vypořádat se dvěma zdroji syntaktické složitosti najednou a jsou tak zahlceny. Tato predikovaná interakce byla experimentálně potvrzena: úspěšnost porozumění u stimulů, které byly z hlediska jazykové syntaxe složitější a zároveň zahrnovaly vybočení z tóniny, byla výrazně nižší než u ostatních jazykově složitých stimulů.

Z výše uvedeného plyne, že interakce mezi percepcí hudby a řeči jsou složité a často navíc zastíněné tím, že jsme pro dekodování řeči až příliš dobře vybaveni. Mnohé úlohy tak vedou ke stropovému efektu a ústí v příliš jednoduchá data, jež obsahují pouze náhodnou variabilitu, která nekoreluje se sledovanou hudební proměnnou. Jistý způsob, jak se tomuto efektu vyhnout, spočívá v tom, že percepční výzkum povedeme na jazyce, který není rodným jazykem posluchačského vzorku, případně zvolíme „nepřirozené“ jazykové úlohy. V takových případech nám pak vztah mezi řečovou a hudební percepcí vyvstává relativně jasně. Vracíme se tedy obloukem k Patelově **SSCLMH** (hypotéze o sdíleném mechanismu pro osvojování zvukových kategorií): hudební vlohy budou patrně klíčové pro osvojení důležitých akustických kontrastů, nicméně samotné výsledné reprezentace budou v mozku uloženy odděleně (viz též Patelova **SSIRH**). Pokud je při percepci vjem spárován s již dříve existující reprezentací, naprosto to pochopitelně přebije výsledky jakékoli paralelně probíhající on-line analýzy; její vliv se může projevit až druhotně při velkém počtu opakování tím, že se díky ní začne přetvářet právě systém reprezentací. Pokud je ovšem ono spárování narušeno (ať už tím, že je vjem samotný pro posluchače nezvyklý, nebo že vyžadujeme jeho neintuitivní rozbor), mělo by být možné sledovat on-line analýzu přímo a dobrat se toho, jak a zda vůbec je v hudbě a řeči propojená.

A composer is a guy who goes around forcing his will on unsuspecting air molecules, often with the assistance of unsuspecting musicians.

Frank Zappa, *The Real Frank Zappa Book*

Část II

Experiment

3 Výzkumné paradigma

V závěru předchozí části jsme dospěli ke zjištění, že lidský mozek je zvláště na percepci řeči velmi dobře trénovaný, a pokud se mu předloží úlohy, na něž je zvyklý, povede to velmi rychle ke stropovému efektu. Jinými slovy, kapitální (*top-down*) procesy převálcují procesy bazální (*bottom-up*).³³ Zmínili jsme též dva způsoby, jak se stropům vyhnout: testovat percepci cizího jazyka, nebo použít „nepřirozené“ úlohy. Je ještě třetí možnost: pečlivě zvolit úroveň nápadnosti změn, u nichž sledujeme, zda si jich posluchač povšimne. Odhadnout tuto úroveň není triviální, jak jsme viděli v odd. 2.2.3, právě postřehnutelný rozdíl a tedy i nápadnost frekvenčních a temporálních změn silně záleží na charakteru stimulů i kontextu. Není tedy moudré se při tvorbě stimulů slepě spolehnout na jednu z uváděných hodnot, naopak je třeba si empiricky ověřit jejich složitost důkladnou předběžnou pilotáží, neboť při příliš těžkých stimulech hrozí zase efekt podlahový (viz Lee a Lee 2010 a popis jejich experimentu v odd. 2.2.4).

S vědomím těchto úskalí jsme zvolili výzkumné paradigma percepčního test, který mapuje percepční citlivost respondentů na jemné frekvenční a temporální variace v hudebních a řečových stimulech. Byly zařazeny tři typy úloh, blíže popsané v kap. 4:

1. úlohy čistě řečové (odd. 4.1)
2. úlohy čistě hudební (odd. 4.2)
3. úlohy, které vyžadují zapojení obou domén (odd. 4.3)

Všechny úlohy testovaly diskriminaci (schopnost porovnat referenční a neznámý zvuk a rozhodnout, zda se liší), ne identifikaci (schopnost zařadit zvuk do předem osvojené kategorie). Respondenti, kteří mají s hudbou malou zkušenost, tak nebyli v tomto ohledu samotným formátem testu nijak znevýhodněni. Naše pracovní hypotéza byla korelačního charakteru: předpokládali jsme, že úspěšnost jednotlivých respondentů ve všech třech typech úloh bude provázaná, tj. kdo bude citlivý na variace v řeči, bude přesněji vnímat i změny v hudbě. Takový výsledek by nepřímo ukazoval k tomu, že řeč i hudba jsou skutečně v percepci jistým způsobem propojené, a toto propojení se může projevit i u dospělých jedinců navzdory silným kapitálním procesům založeným na doménově specifických mentálních reprezentacích. Kromě toho jsme u respondentů sledovali i další faktory, které by na výkon v testu mohly mít vliv: jejich obecný vztah k hudbě a případné vysokoškolské vzdělání zaměřené na jazyk.

³³Kapitální procesy jsou takové, které při percepci využívají struktur koncového mozku k predikci na základě předchozí zkušenosti; bazální jsou naopak ty, které se zakládají pouze na bezprostředním jevu. Metafora *top-down* naznačuje, že vjemu „seshora“ nutíme interpretaci, s níž jsme již obeznámeni, kdežto *bottom-up* znamená, že začínáme analýzou samotného vjemu a jeho interpretaci stavíme „odspodu“. Viz též obr. 3 na str. 3, kde „kapitální“ odpovídá šipkám „statistická predikce“ a „bazální“ šipkám „lokální vyvozování“.

Zdůrazněme ještě, že na rozdíl od výzkumů pomocí fMRI, ERP apod. popsaných v odd. 2.2.1, které sice přinášejí zajímavé výsledky, ale z pochopitelných důvodů pro nás v rámci omezených prostředků nebyly možností, nezkoumá percepční test neurofyzilogické či přímo kauzální propojení mezi oběma doménami, nýbrž pouze korelace ve vědomém chování subjektů.

4 Metoda

Pro percepční test byly sestaveny tři sady stimulů: 18 řečových (věty), 18 hudebních (melodie) a 18 hudebních protějšků (analogů) vytvořených jako hudební stylizace stimulů z první sady (detaily níže). Ke každému stimulu byly vytvořeny melodicky a rytmicky manipulované obměny, čímž vznikly položky o třech zvucích (vždy dvakrát původní stimul a jednou jedna z jeho zmanipulovaných verzí), u nichž bylo úkolem rozpoznat, které dva zvuky jsou stejné.³⁴ Položky byly nejprve pilotně odzkoušeny na malém počtu respondentů a posléze z nich byl sestaven test o třech částech (řečové, hudební a smíšené) a celkově 112 položkách (40 + 40 + 32). Testu se zúčastnilo 42 respondentů.

4.1 Řečové stimuly

Řečové stimuly spočívaly v 18 krátkých větách, z nichž každá byla tvořena třemi mluvnickými taktů: dvěma trojslabičnými a posledním dvojslabičným. Pro větší rozmanitost (aby se neztrácel jazykový charakter stimulů a respondenti je nepřestali vnímat jako řeč) byly využity dva větné typy, věta oznamovací a otázka zjišťovací. V každé větě byl první vokál jednoho ze tří taktů vybrán jako cílový pro další manipulaci se stimuly (viz níže). Aby se předešlo nechtěné variabilitě ve stimulech způsobené mikrintonačními jevy, byly veškeré cílové vokály [a] nebo [a:]. Tato hláska byla zvolena proto, že je ve tvarech českých slov hojně zastoupena v krátké i dlouhé variantě, jež se navíc na rozdíl např. od i-ových hlásek liší opravdu pouze trváním, a jako nejotevřenější český vokál je také nejsonornější, tudíž se na ní dobře sleduje průběh intonace.

Délka se střídala z toho důvodu, že na cílovém vokálu byly dále prováděny i temporální manipulace (zkracování a prodlužování). Proto jsme sledovali také délku vokálu následujícího po cílovém, která může mít vliv na stanovení referenčního rámce pro trvání vokálu uvnitř mluvnického taktu, a aby si byly cílové taktů svou vokalickou strukturou co nejpodobnější, drželi jsme se požadavku, aby i případná třetí slabika měla za vokál [a].³⁵ Další jevy, které mohou mít na percepci rytmu v rámci taktu vliv, jsme již rigorózně nesledovali: např. co se slabičné struktury týče, převažují v cílových taktech otevřené slabiky, nikoli ovšem výhradně. Je otázkou, zda v další práci neošetřit i tento aspekt, či se přímo nezaměřit na zkoumání citlivosti na změny trvání vokálů v závislosti na stavbě taktu (délky vokálů, délka slabičné préturey či přítomnost nebo absence kody).

Celou paletu kombinací čtyř výše popsaných parametrů nebylo možné v 18 větách

³⁴V části testu využívající hudební analogy byl úkol trochu jiný: rozpoznat, který ze dvou zvuků z jedné domény (např. řeč) je bližší zvuku z domény druhé (např. hudba). Blíže viz odd. 4.3.

³⁵Proto nebyla mezi použité větné typy zahrnuta otázka doplňovací, neboť ta vyžaduje na první pozici ve větě tázací slovo, které znemožňuje vyhovět těmto požadavkům. Vhodní kandidáti na pozici cílových taktů s příslušnou vokalickou strukturou byli vyhledáni v korpusu SYN2010 (ÚČNK FF UK 2010) pomocí následujícího dotazu jazyka CQL: "[^ěéýýííoóuúáá]*[aá]{2,3}[^ěéýýííoóuúáá]*".

pokryt (2 větné typy \times 3 umístění cílového taktu \times 2 délky cílového vokálu \times 2 délky následného vokálu = 24) a navýšení počtu vět se nezdálo přínosné, i s ohledem na únosnou délku pilotáže. Nicméně jsme se snažili o pestré zastoupení různých variant i s vědomím toho, že manipulace na některých pozicích v kombinaci s jistým větným typem či délkou cílového nebo následujícího vokálu mohou být nápadnější než jiné.

věta	větný typ	cílový MT	cílový V	následný V
Párátka / donesou / taky?	ot. zjišť.	1	a:	a:
Vstávala / Andulka / brzy?	ot. zjišť.	1	a:	a
Mávátka / dodali / včera.	ozn.	1	a:	a:
Nemilá / závada, / zdá se.	ozn.	2	a:	a
Byla tu / spáchána / vražda?	ot. zjišť.	2	a:	a:
Viděl jsem / mláďata / losa.	ozn.	2	a:	a
Josefa / svíčková / láká?	ot. zjišť.	3	a:	a:
Výdaje / pokryje / vlada?	ot. zjišť.	3	a:	a
Je to jen / Petrova / známá.	ozn.	3	a:	a:
Začala / poslední / sada.	ozn.	1	a	a
Zahrála / na svatbě / tango?	ot. zjišť.	1	a	a:
Kanada / vyhrála / hokej.	ozn.	1	a	a
Pronikla / armáda / do bran?	ot. zjišť.	2	a	a:
Budou ta / rajčata / čerstvá?	ot. zjišť.	2	a	a
Tahleta / kavárna / vede.	ozn.	2	a	a:
Rusové / zamítli / radar.	ozn.	3	a	a
Anežka / byla ta / malá?	ot. zjišť.	3	a	a:
Veleboj / titánů / začal.	ozn.	3	a	a

Tabulka 1: Věty použité pro řečové stimuly, jejich odpovídající větné typy, pořadí cílového mluvního taktu ve větě a typ cílového a následného vokálu ($a \times a:$, tj. krátký \times dlouhý). Lomítka v prvním sloupci označují hranice mluvních taktů.

Všechny věty namluvil sám autor, se slovním přízvukem vždy na první slabice každého

taktu, v tiché místnosti na kvalitní přenosný diktafon Sony ICD-UX522 (formát WAV – Microsoft PCM, bitová hloubka 16 bitů, vzorkovací frekvence 44,1 kHz). Lehký šum v pozadí byl odstraněn pomocí příslušné funkce v programu Audacity, která šum potlačuje na základě jeho předem získaného profilu v rámci nahrávky. Kvalita nahrávek tedy nebyla studiová, ale vzhledem k tomu, že záměrem nebylo testovat absolutní vlastnosti lidského sluchu, nýbrž relativní (citlivost na rozdíl), se domníváme, že výsledky studie tím nejsou nijak znehodnoceny.³⁶

Jednotlivé věty byly v programu Praat (Boersma a Weenink 2014) nejprve převedeny na objekty typu *Manipulation* (se standardními nastaveními: časový krok 0,01 s, minimální frekvence 75 Hz, maximální frekvence 600 Hz), v nichž lze metodou PSOLA manipulovat s trváním a základní frekvencí vybraných úseků zvuku, a to nezávisle na sobě. Poté bylo provedeno celkové vyhlazení křivky F0 (tzv. stylizace, s prahem 0,5 ST) a věty s intonací resyntetizovanou podle této stylizace byly uloženy namísto původních nahrávek jako referenční, nemanipulované stimuly. Důvodem pro toto rozhodnutí bylo, aby jak referenční, tak manipulované stimuly prošly jistou umělou úpravou a resyntézou, aby se posluchači nemohli rozhodovat na základě případných malých rozdílů v přirozenosti obou zvuků. Zároveň se tak usnadnila manipulace s F0: stylizací se redukuje počet výškových bodů, které konturu charakterizují (zbytek se lineárně interpoluje), a jejich lokální posuny tedy místo „zubů“ v kontuře ústí v povlnnější přechody.

Na cílovém vokálu každé věty pak byly systematicky provedeny dva druhy manipulací v několika úrovních nápadnosti: manipulace **melodická** (celkový posun průběhu intonace v cílovém místě níže či výš) a manipulace **rytmická/temporální** (zkrácení či prodloužení trvání cílového vokálu). Melodická manipulace spočívala ve skriptu, který posunem všech výškových bodů v rámci cílového vokálu vygeneroval intonační varianty v rozpětí <-4 ST; +4 ST> v půltónových krocích.³⁷ Skript pro rytmickou manipulaci pak generoval varianty, v nichž byly krátké cílové vokály dloouženy a dlouhé naopak kráceny, a to o 0,2, 0,3, 0,4 nebo 0,5-násobek svého původního trvání.

Na základě poslechu a impresionistického posouzení pak autor vybral z různých možností směru a úrovně manipulace pro každou větu jednu melodickou a jednu rytmickou manipulaci, z nichž pak v kombinaci s původní verzí vytvořil položky, které zařadil do pilotáže. U některých rytmických variant s dlooužením přitom sáhl k ještě výraznějšímu prodloužení (až na dvojnásobek původní délky), než bylo maximum v původní sadě, neboť změna se nezdála být dostatečně výrazná.

³⁶Maximálně se tento fakt může projevit v plošně o něco horších výsledcích.

³⁷Pokud po provedení stylizace nezůstal v hranicích vokálu žádný výškový bod (tj. zbyla jen interpolace), byl posunut bod nejbližší hranicím vokálu, případně dva, pokud se takoví kandidáti naskýтали z obou stran.

4.2 Hudební stimuly

Zatímco řečové stimuly byly tvořené třemi mluvními takty, skládaly se stimuly hudební ze tří 2/4 taktů hudebních. Není přitom záměrem tvrdit, že si tyto jednotky navzájem odpovídají; že nesou stejné jméno, vyplývá z metaforického přenesení pojmu „takt“ z oblasti hudební teorie do jazykovědy, a je tedy podobností intuitivní spíš než formální. Zatímco mluvní takt je seskupení prozodicky svázaných slabik, tóny v hudebním taktu nutně všechny provázané být nemusí, a naopak: vjemově „těsné“ seskupení tónů může jednoduše hranici taktu překročit. Protějškem mluvního taktu je v hudbě mnohem spíš jednotka na úrovni ucelené podsekvence hudebního motivu.

Paralelní použití tří taktů má zde tedy prozaičtější důvod: zaprvé, zajistit u všech melodií srovnatelné trvání, a zadruhé, podobně jako u řečových stimulů nabídnout tři různá umístění pro rytmické a melodické variace, přímočaře ukotvená vždy na začátku prvního, prostředního a posledního taktu. Podkladový dvoučtvrťový rytmus přitom nebyl ve zvuku nijak zhmotněn (žádný metronom či perkuse) a melodie se výrazně lišily v tom, jak moc se mu poddávaly (tj. dávaly ho znát) či naopak šly proti němu a vytvářely polyrytmické napětí. Rozpětí v typech melodií v tomto ohledu sahalo od prostého trojího zopakování stejného motivu, kde byly rytmus i hranice taktů zřetelné, až po útvary, kde bez rytmického podkladu zněl tón spadající na první dobu spíše předražně. Některé melodie byly víceméně uzavřené fráze, některé byly naopak cíleně navrženy tak, aby působily vytrženě z kontextu.

Celkem vzniklo 18 krátkých melodií, dvě od každého z devíti typů definovaných kombinací dvou parametrů: zda melodie mezi cílovým a následujícím tónem stoupá, klesá, či zůstává na stejné úrovni, a zda je cílový tón oproti následujícímu kratší, delší či stejně dlouhý. Předpokládali jsme totiž, že právě tyto aspekty mohou mít velký vliv na výslednou nápadnost posléze provedených manipulací. Kromě toho se pro větší rozmanitost v melodiích střídaly i následující charakteristiky:

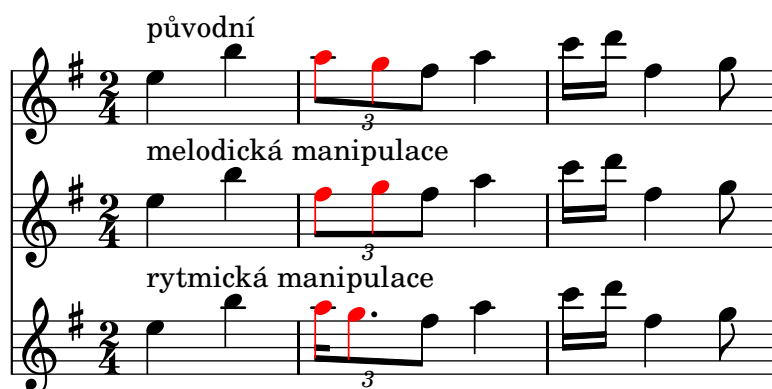
1. 6 modů: jónský, dórský, frygický, lydický, mixolydický, lokrický
2. od 3 různých tónik: H, C, D
3. na 9 nástrojů: Hammondovy varhany, viola, trubka, hoboj, příčná flétna, klavír, marimba, cembalo a smyčcové pizzicato³⁸

Většina melodií tedy vycházela z jiného církevního modu než klasické durové (jónské) stupnice, která je současným západním uším dobře známá a odchylky od ní při melodických manipulacích jsou tedy potenciálně nápadnější. Povšimněte si také, že použité nástroje se dělí na skupinu s neomezeným *sustainem* (prvních pět nástrojů v seznamu

³⁸Odpovídající názvy a identifikační čísla ve standardu MIDI, který byl použit při generování zvuku: Rock Organ (18), Viola (41), Trumpet (56), Oboe (68), Flute (73), Acoustic Grand Piano (0), Marimba (12), Harpsichord (6), Pizzicato Strings (45).

výše), u nichž lze držet stabilní fázi tónu v principu donekonečna, a skupinu se *sustainem* omezeným (zbývající čtyři nástroje v seznamu výše), u nichž stabilní fáze tónu dříve či později nutně dozní. To může mít vliv zejména na vnímání rytmických/temporálních manipulací, neboť zatímco u neomezeného *sustainu* zkracujeme či prodlužujeme trvání plného tónu, u omezeného takto měníme pouze trvání jeho úpadku.

Melodie byly sestaveny s pomocí knihovny `leipzig`³⁹ v jazyce Clojure, která umožňuje jak pohodlně předem zvolit tónový materiál dané stupnice a nadále se pohybovat pouze v něm, tak jednoduše manipulovat se symbolickým zápisem melodie. V této fázi byly tedy rovněž rozgenerovány melodické a rytmické varianty původních stimulů. U změn melodie šlo znovu o rozpětí $<-4 \text{ ST}; +4 \text{ ST}>$ v půltónových krocích, u rytmu o zkrácení či prodloužení (podle toho, zda byl cílový tón delší nebo kratší než následující) o zlomek původního trvání cílového tónu ($1/2$, $1/3$, $1/4$, $1/6$ nebo $1/8$). Na rozdíl od řečových stimulů byl u melodií také zrcadlově prodloužen či zkrácen i tón následující po cílovém, aby byla změna skutečně jen lokální a zbytek melodie nebyl vůči metru nijak posunut. Pomocí knihovny `MIDIUtil`⁴⁰ v jazyce Python byly pak interní symbolické zápisy melodií a jejich variant převedeny na standardizovaný symbolický zápis binárního formátu MIDI, který byl posléze programem `TiMidity++` renderován již do zvukového formátu WAV (Microsoft PCM, bitová hloubka 16 bitů, vzorkovací frekvence 44,1 kHz). Vzorky nástrojů použité pro „přehrání“ MIDI souborů pocházely ze soundfontu `FluidR3_GM.sf2`. Příklad melodie a jejích pozměněných protějšků z finální verze testu naleznete v obr. 7.



Obrázek 7: Příklad melodie a jejích melodicky a rytmicky manipulovaných verzí, které byly použity ve finálním testu. Stupnice je C lydickej, použitý nástroj: příčná flétna.

Konkrétní hodnoty manipulací použité v pilotáži vybral sám autor podobným způsobem jako u řečových stimulů, přičemž přihlížel k tomu, aby u melodických manipulací byly pokud možno rovnoměrně zastoupeny tyto typy případů:

1. manipulací dochází k porušení kontury tak, že mezi cílovým a následným tónem

³⁹Dostupné na <https://github.com/ctford/leipzig>.

⁴⁰Dostupné na <https://code.google.com/p/midiutil/>.

- klesá místo aby stoupala
- manipulací dochází k porušení kontury tak, že mezi cílovým a následným tónem stoupá místo aby klesala
 - manipulací k porušení kontury nedochází

4.3 Hudební stimuly pro smíšené položky

U smíšených položek šlo o vytvoření hudebních analogů k výše zmíněným 18 větám, které jim budou melodicky a rytmicky odpovídat, jako by se daný hudební nástroj snažil mluvčího imitovat. Účelem byla tedy podobnost čistě po zvukové stránce, ne snaha o vyjádření větného obsahu.

Tato úloha nemá triviální, jednoznačné řešení; ani nalezená literatura k ní nenabízela příliš mnoho vodítek, snad kromě domněnky Lehistové o percepčním zpravidelnění řečového rytmu (viz str. 16). Na jednu stranu se nabízela možnost věrně temporálně i melodicky okopírovat (tedy na hudební nástroj „přehrát“) průběh prozogramů. Stimuly by ovšem v takovém případě ztratily na hudebním charakteru, přičemž cílem v této práci bylo přece jen zachovat nějaký tonální a temporální rámec a s ním jistou konvenční hudebnost. Navíc by takový postup kvůli atonálnímu charakteru intervalů vyžadoval složitější technické řešení při generování hudební melodie než běžně dostupné MIDI.

Po různých pokusech jsme se nakonec ustálili na následujících pravidlech pro tvorbu hudebních protějšků:

- každá slabika je reprezentována jedním samostatným tónem
- trvání tónu se odvozuje od délky vokálu (složitost slabičné kody tedy nehraje primárně roli), což umožňuje jednoduše převést temporální vztah mezi cílovým a následným vokálem: krátké vokály buď $1/4$ nebo $1/3$ doby, dlouhé vokály $2/5$ nebo $1/2$ (zvlášť v posledním mluvnickém taktu, kde lze předpokládat koncové zpomalování)
- výška tónu odpovídá řeči pouze v relativním slova smyslu, tj. intervaly mezi sousedními tóny jsou založeny na rozdílech v půltónech mezi úrovněmi intonační křivky ve dvou sousedních vokálech, zaokrouhlených na celé půltóny; jinak se melodie pohybuje v rámci dvoučárkované oktávy a přilehlém okolí

Postup přitom nebyl plně mechanický, na to jsou výše popsána pravidla příliš heuristická, a podoba hudebního analogu byla tedy průběžně upravována podle percepčního dojmu ze srovnání s řečovým originálem (nejprve čistě rytmická složka, pak i ve spojení s melodií).

Úhrnem došlo tedy k melodickému a temporálnímu zpravidelnění, byť v tom druhém ohledu jen částečnému: jak je ze způsobu volby trvání tónu patrné, noty odpovídající mluvnickému taktu nemusejí časově přesně spadat vždy do izochronních hudebních taktů.

Nicméně i tato temporální rozvolněnost má v hudbě obdobu (výrazové prostředky jako *ritenuto*, *rubato* aj.), takže do konvenčního hudebního jazyka patří, není nesrozumitelná či přespříliš rušivá.

Vznikly tak hudební analogy, které možná nejsou „ideální“ nebo „přesné“ (pokud něco takového existuje), neboť nevycházejí z empirického testování většího počtu konkurenčních variant převodu řeči na hudbu na reprezentativním posluchačském vzorku. Nicméně jsou rozhodně tvořeny konzistentně na základě řečových stimulů, takže zdatnější posluchač v nich paralely po jisté nezbytné akomodaci objeví, což se v praxi potvrdilo. Zvuky byly generovány podobným způsobem jako u čistě hudebních položek, jen manipulované varianty byly tvořeny rovnou cíleně tak, aby manipulace byla nápadná, neb u smíšených položek nebyl úkol posluchačů rozdíl mezi zvuky jen identifikovat, nýbrž též rozhodnout, která z variant je zvukově blíže referenčnímu stimulu z opačné domény. Do pilotáže šlo 18 položek, rovnoměrně rozdělených mezi případy, kde byla referenčním stimulem řeč nebo hudební analog, a taktéž podle druhu manipulace (melodie či rytmus).

4.4 Pilotáž a sestavení testu

Pilotáž proběhla ve třech oddělených fázích podle typu stimulů (jinak by byla délka testu neúnosně dlouhá) za účasti dvou a u řečových stimulů tří respondentů. Ti kromě podstoupení testu poskytli ještě slovní zpětnou vazbu.

V řečové a hudební části bylo pořadí stimulů v rámci položek AXB, tj. referenční stimul zazněl uprostřed, a úkolem bylo poznat, který z obou krajních stimulů je stejný jako prostřední (neobsahuje melodickou nebo rytmickou manipulaci). Řazení AXB představuje způsob, jak dostat oba neznámé stimuly těsně ke stimulu referenčnímu, aby jeden z nich nebyl svou bližší pozicí zvýhodněn; s podobným účelem je použito např. ve studiích Brancazio et al. (2006) nebo Kroos a Hogan (2009). V položkách smíšené části bylo pořadí stimulů XAB, neboť ve chvíli, kdy referenční zvuk pochází z jiné domény, je přirozenější nejprve se ukotvit na originálu a teprve posléze posuzovat věrnost „imitací“. Úkolem bylo určit, která z imitací je přiléhavější a originální zvuk lépe napodobuje. Neznámé stimuly se vždy objevily v obou možných pořadích, tj. AXB i BXA.

Mezi položkami se pokaždé ozval krátký desenzitační zvuk. Respondenti měli možnost si nechat každou položku jednou zopakovat a po zadání odpovědi měli ještě ohodnotit subjektivní stupeň jistoty na škále „jsem si úplně jistý/á“, „jsem si celkem jistý/á“ a „spíš hádám“.⁴¹ Pilotáž i následný ostrý test byly implementovány v prostředí programu Praat pro zadávání percepčních testů.

Na základě výsledků pilotáže bylo vybráno 10 hudebních, 10 řečových a 8 smíšených

⁴¹První dva stupně pokrývají případy, kdy má respondent pocit, že se rozhodoval nějakým způsobem fundovaně (ne čistě náhodně).

položek, vždy půl napůl s melodickou a rytmickou manipulací. Celkem měl tedy finální test 28×2 pořadí neznámých stimulů $\times 2$ opakování = 112 položek rozdělených do tří bloků podle domény (řečová, hudební, smíšená). Dva stimuly v řečové i hudební části byly schválně zařazeny s oběma druhy manipulace (tj. stejný referenční zvuk, ale jednou měl distraktor změněný rytmus a podruhé melodii). Výběr položek směřoval k pokrytí široké škály náročnosti v rámci každé kategorie, od těch, které byly pro respondenty jednoduché, až po ty složitější. U některých byla ještě dodatečně manipulace zvýrazněna, protože celkově se úlohy ukázaly jako poměrně tvrdý oříšek.

Ve smíšeném bloku byly nakonec zařazeny pouze položky, kde byl referenční stimulus řečový a neznámé stimuly (protějšky) hudební, byť pilotáž zahrnovala i opačný směr.⁴² Ukázalo se totiž, že přeladit na jiný typ referenčního stimulu v rámci bloku není jednoduché, a na další samostatný blok už nebyl prostor (test by byl příliš dlouhý). Z obou možných směrů se pak jako intuitivnější ukázal pokyn pokusit se vnímat melodii jako imitaci řeči než ten opačný.⁴³

Pro lepší představu o rozpětí manipulací v položkách finálně začleněných do ostrého testu nabízíme jejich přehled v tab. 2, byť je dobré mít na paměti, že jejich výsledná nápadnost nevyplývá jen paradigmaticky ze stupně manipulace, nýbrž i syntagmaticky z trvání a výšky okolních jednotek. Hodnoty u smíšených stimulů, kde byl manipulovaný analog vědomě tvořen tak, aby se pokud možno jasně lišil od správné odpovědi, v tabulce neuvádíme.

řeč – melodie	řeč – rytmus	hudba – melodie	hudba – rytmus
-4 ST	0.7×	-3 ST	0.67×
-4 ST	0.8×	-1 ST	0.75×
-3 ST	1.4×	+1 ST	0.83×
+3 ST	1.4×	+2 ST	1.33×
+3 ST	2×	+3 ST	1.5×

Tabulka 2: Hodnoty manipulací řečových a hudebních stimulů vybrané do finálního testu. U melodických manipulací se jedná o posun nahoru či dolů v čtvrttónech, u rytmických o násobek trvání původního tónu či vokálu.

⁴²Referenčním stimulem byl tedy hudební analog popsáný v odd. 4.3 a byl porovnáván s původní podobou věty, podle níž byl vytvořen, a nějakou její manipulovanou verzí.

⁴³Je asi přirozenější, aby se procesem imitace část informace (v našem případě segmentální obsah vět) ztratila, než aby byla naopak volně doplněna.

4.5 Ostré testování

Pro co největší homogenitu v proměnných, které by mohly mít na výsledky rušivý vliv, bylo všech 42 respondentů vybráno z populace studentů VŠ ve věkovém rozmezí od 18 do 30 let. Všichni byli rodilými mluvčími češtiny a projít testem jim trvalo přibližně od 30 minut až po 1 hodinu. Testování se odehrávalo ve zvukově izolovaném studiu na Fonetickém ústavu FF UK, za použití tichého přenosného počítače a stejného páru sluchátek pro všechny účastníky.

Na začátku každé ze tří částí se nejprve zobrazily pokyny k úkolu v dané části, načež si respondent úlohu vyzkoušel na dvou záchvičných položkách. Jednotlivé položky byly stejně jako v pilotáži odděleny desenzitačním zvukem, i možnost opakovaného přehrání položky a nutnost zadat po odpovědi i subjektivní stupeň jistoty zůstaly stejné. Testovací prostředí programu Praat je pouštělo v pseudonáhodném pořadí, tak aby se stejná položka neopakovala dvakrát za sebou (páry AXB a BXA byly v tomto ohledu z technických důvodů považovány za dvě různé položky). Vždy na konci a uprostřed každé části vybídl test respondenta, aby si udělal podle svého uvážení pauzu. Respondenti věděli, že mají posuzovat zvukovou podobnost stimulů, ale nic o konkrétním charakteru možných odlišností (manipulace s melodií či rytmem) jim řečeno nebylo.

Po skončení testu vyplnili respondenti krátký dotazník týkající se převážně jejich vztahu k hudbě a praktickým zkušenostem s ní (tab. 3). Jedna z otázek se pak ptala na typ studované či absolvované vysoké školy na škále od nijak speciálně jazykově zaměřené přes filologické obory až po fonetiku (celkem pět kategorií). Cílem bylo odhadnout míru, do níž jsou jednotliví respondenti vědomě obeznámeni s principy zvukové stavby řeči, což mohlo představovat výrazné usnadnění při identifikaci drobných manipulací v řečové signálu, u fonetiků případně i obecně vytríbenější poslechové návyky.

otázka	nabízené odpovědi
Věk:	přirozené číslo
VŠ jazykového zaměření?	“ne”, “mezinárodní vztahy, teritoriální studia apod.”, “filologie”, “lingvistika”, “fonetika”
Hudební vzdělání:	“žádné”, “soukromé lekce”, “ZUŠ”, “konzervatoř”, “VŠ/VOŠ”
Délka hudebního vzdělání (roky):	přirozené číslo
Působení v hudebním tělese (sbor, orchestr, kapela):	“ne”, “ano”
Poslech hudby (kulisa = pasivní, sledování skladby = aktivní):	“převážně pasivní”, “i aktivní”

otázka	nabízené odpovědi
Hudební aktivita v současnosti:	“ne”, “ano”
Tanec:	“vůbec”, “příležitostně”, “pravidelně”, “profesionálně”
Aktivní pravidelná hra (dříve či aktuálně) na:	“ne”, “ano”
<ul style="list-style-type: none"> • strunné nástroje • klávesové nástroje • dechové nástroje • perkusní nástroje 	

Tabulka 3: Dotazník, který respondenti vyplňovali po skončení percepčního testu. Primárně se týkal jejich vztahu k hudbě.

Dotazy týkající se hudby v tab. 3 (všechny kromě prvních dvou) byly integrovány do celkového skóre hodnotícího deklarovanou „hudebnost“ (vztah k hudbě, formální hudební vzdělání a zkušenosti) každého respondenta. Odpovědi byly ohodnoceny v jednotkových krocích od 0 nahoru podle stoupající „hudebnosti“ (viz pořadí v tab. 3) a sečteny.⁴⁴ Respondenti byli podle dosaženého skóre následně rozčleněni do pěti vzestupných kategorií *a–e* tvořených tak, aby v každé byl podobný počet subjektů a zároveň aby ti, kteří mají stejné skóre, byli ve stejné kategorii.

Důvod, pro převod na kategorie je ten, že by bylo přehnané předstírat, že s každým přiděleným bodem nějak rovnoměrně roste respondentova předpokládaná „hudebnost“. Bodo-
vání se snaží (arbitrárním, nicméně konzistentním způsobem) do jednoho skóre zkompile-
vat těžko souměřitelné parametry. Je tedy lepší převést skóre na kategorickou proměnnou,
která respondenty rozdělí do volnějších skupin, než použít přímo číselnou hodnotu. Řeší
se tím i problém možné korelace odpovědí na jednotlivé otázky, která by v extrémním
případě mohla vést k tomu, že by subjekt za stejný aspekt svého vztahu k hudbě dostal
body vícekrát, což by nelinearitou číselné škály ještě zvyšovalo.

Úspěšnost každého respondenta v každé ze tří částí testu pak byla převedena na celkové
skóre podle následujícího vzorce:

$$s = \sum_{o \in O} o_{správně?} \times o_{jistota}$$

⁴⁴Délka hudebního vzdělání byla pro tento účel přetvořena na kategorickou proměnnou.

kde:

- s je výsledné skóre
- O je množina otázek v dané části testu a o odpověď jednu z nich
- $o_{správně?}$ je 1 pokud je odpověď správně a -1 pokud je špatně
- $o_{jistota}$ je 1, pokud respondent uvedl, že si byl „úplně“ či „celkem“ jistý, a 0 pokud „spíš hádal“

Skóre se tedy pohybovala v rozmezí $< -40; 40 >$ v hudební a řečové části a $< -32; 32 >$ v části smíšené, přičemž spodní hranice odpovídá případu, kdy se respondent pokaždé spletl a zároveň souvisle tvrdil, že si je jist, a horní hranice případu, kde si byl pokaždé jist a zároveň odpovídal správně. Mezi oběma úrovněmi jistoty jsme při stanovování skóre nečinili rozdíl z toho důvodu, že je tak možné částečně kompenzovat rozdíly v sebevědomí mezi respondenty (někdo se může „stydět“ tvrdit, že si je „úplně“ jistý). Navíc nám šlo primárně o vyčlenění odpovědí, kde sami respondenti přiznali, že se rozhodovali úplně náhodně.

Závěrem dodejme, že test samotný (včetně zdrojového kódu pro hudební stimuly) a výsledky všech 42 respondentů (v dále nezpracované podobě, ve formátu jeden soubor `.csv` na respondenta), jsou k dispozici ke stažení na adrese <https://github.com/dafydd-lukes/diplomka> Kdo by s nimi chtěl dále pracovat, necht' se prosím ozve autorovi, bude mu poskytnut základní orientační komentář.

5 Výsledky

5.1 Charakteristika respondentů

Nejprve se stručně podívejme na charakteristiku respondentů, kteří se testu zúčastnili, na základě výsledků dotazníku. Obr. 8⁴⁵ ukazuje, že většina subjektů pocházel zprostředka věkového rozmezí 18–30 let. Oproti tomu délka hudebního vzdělání byla poměrně rozmanitá, byť převažoval údaj do 5 let; nemalá část respondentů (16) uvedla délku nulovou.

Jak je vidět z obr. 9, počet respondentů, kteří se v současné době věnují hudbě či někdy působili v hudebním tělese je méně než poloviční. Zastoupení hudebních nástrojů je celkem rovnoměrně rozděleno mezi strunné, klávesové a dechové nástroje, čistě rytmické perkuse poněkud pokulhávají.

Navzdory poměrně nízkému počtu aktuálně aktivních hudebníků obr. 10 ukazuje, že více než polovina respondentů někdy navštěvovala ZUŠ nebo brala soukromé lekce hudby; nadpoloviční je i poměr těch, kteří hudbu poslouchají aktivně (všimají si variací v motivech, změn rytmu či tempa, modulací apod.). Tanci se většina věnuje alespoň příležitostně, byť někteří (13) vůbec. Co se týče absolvovaných či rozstudovaných VŠ, nejvíce respondentů zvolilo obory, které nejsou nijak specificky jazykově zaměřené. Následují filologové a fonetici, naopak studenti mezinárodních vztahů či obecné lingvistiky tvoří jen okrajové skupiny.

5.2 Konzistence a správnost odpovědí

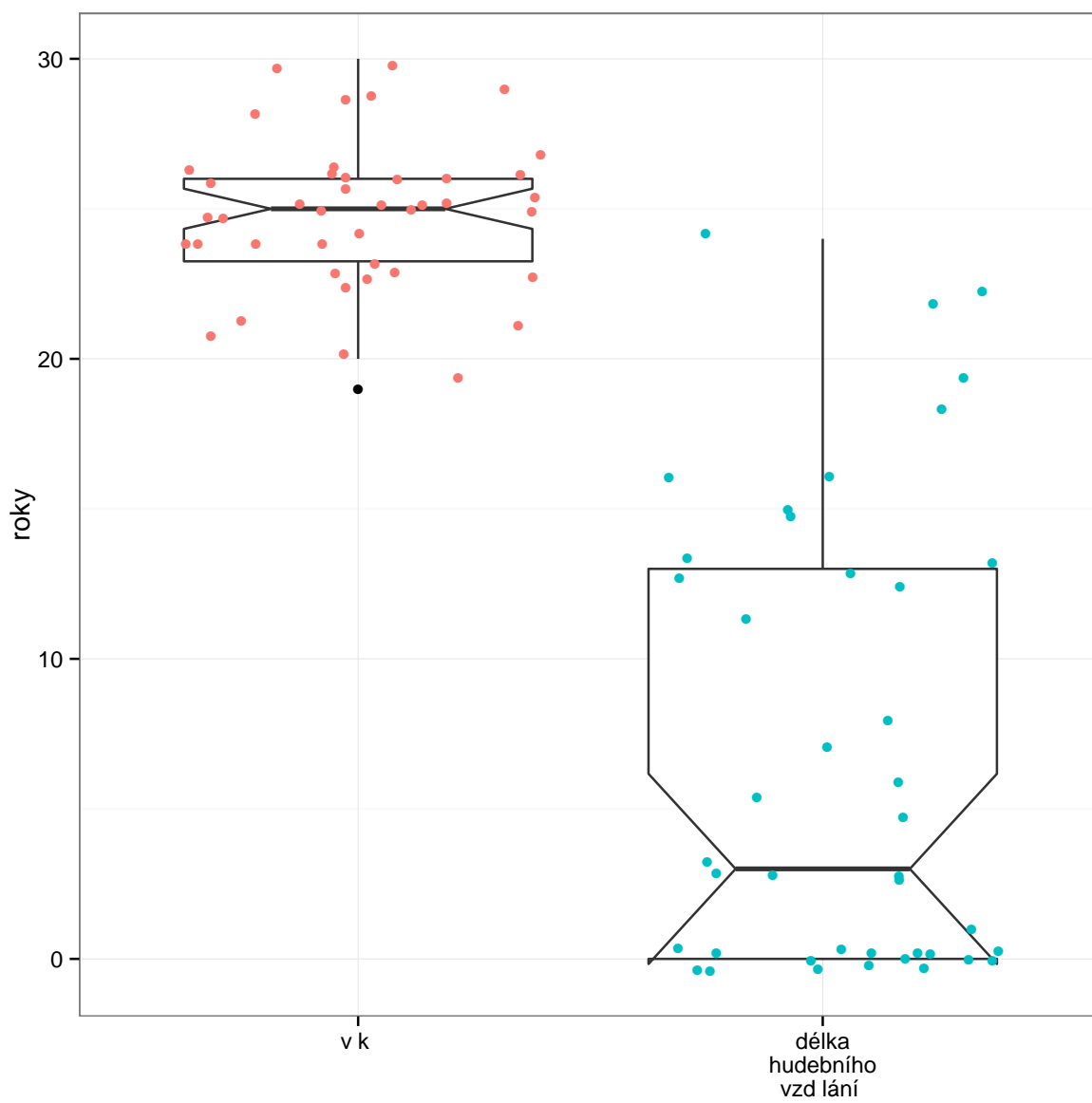
Snažit se hodnotit konzistenci odpovědí na jednotlivé položky napříč subjekty⁴⁶ zde není užitečné: takové míry mají smysl ve chvíli, kdy jsou hodnotící kategorie subjektivní a je třeba ověřit intersubjektivní shodu. V našem případě byla naopak správná řešení úloh vždy objektivně stanovitelná⁴⁷ a test byl cíleně sestaven tak, aby alespoň pro část subjektů bylo těžké tuto správnou odpověď rozpoznat a šlo tak zkoumat korelaci mezi úspěšnostmi v jednotlivých částech testu.

Zajímavější je vyhodnotit konzistenci uvnitř subjektů, tj. zda na stejnou položku volili

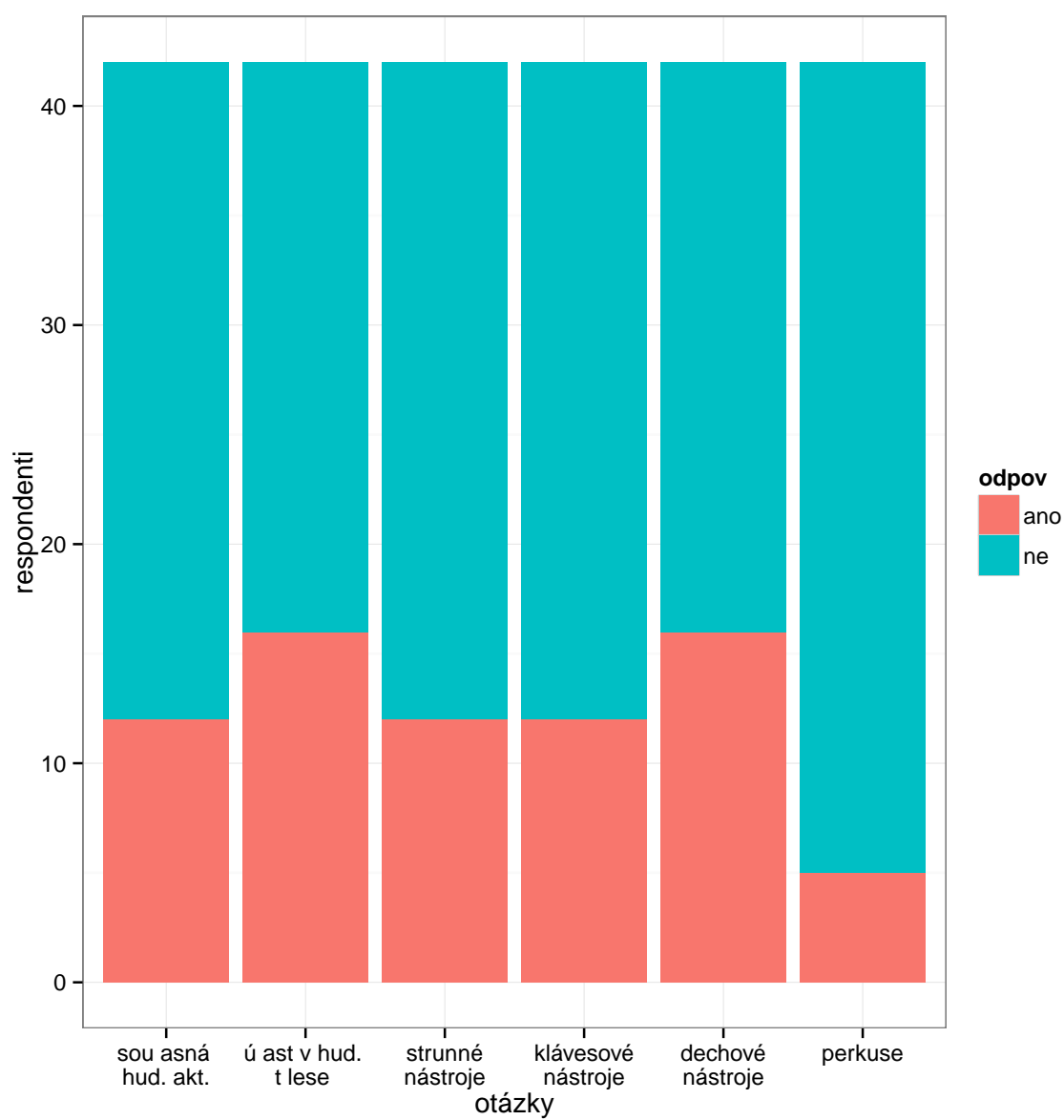
⁴⁵Pro tento a všechny následující krabicové grafy platí, že **vodorovná čára** představuje medián, **krabice** mezikvartilové rozpětí (IQR), **vruby na krabicích** $\frac{1,58 \times \text{IQR}}{\sqrt{N}}$ (přibližné rozpětí pro 95% interval spolehlivosti mediánu), **vertikální čáry** sahají k nejvyšší/nejnižší hodnotě v rámci $1,5 \times \text{IQR}$ od kraje krabice, a **černobílé body** představují odlehlé hodnoty za touto hranicí. Barevné body pak odpovídají všem pozorovaným hodnotám v dané kategorii a jsou umístěny s jistou dávkou náhodného šumu, aby se nepřekrývaly.

⁴⁶Např. pomocí Fleissovy κ , která je generalizací Cohenovy κ pro více než dva respondenty.

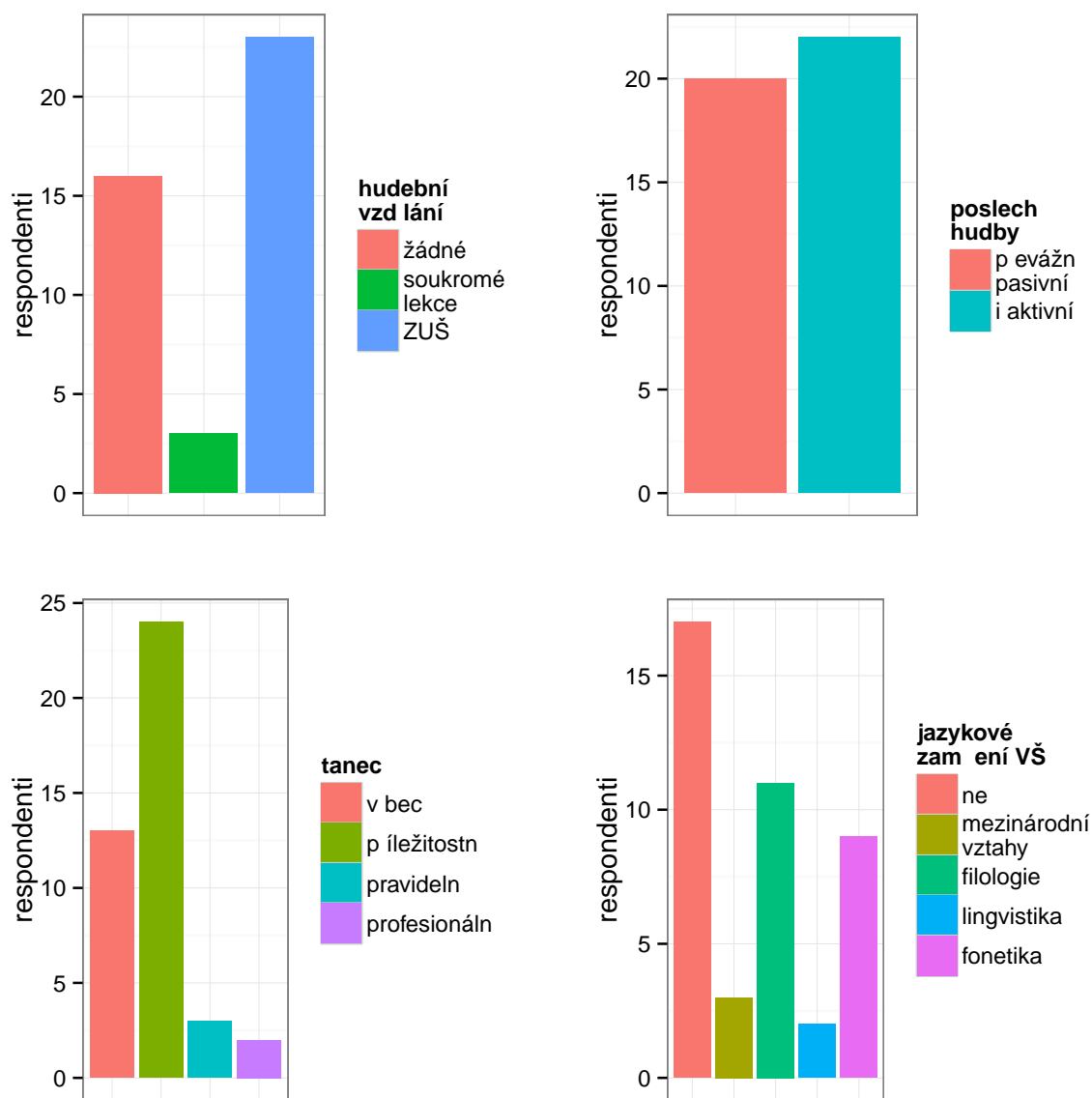
⁴⁷Ve třetí, smíšené části toto tak úplně neplatí, neboť je otázkou, jaké přesně by měl hudební analog mít parametry, aby co nejpřiléhavěji představoval svůj řečový protějšek. V našem případě jsme se z praktických důvodů řídili jen hrubými výsledky pilotáže a intuicí, ale problematika by stála za vlastním výzkum.



Obrázek 8: Zastoupení odpovědí na otázky z dotazníku v tab. 3 na škále přirozených čísel.



Obrázek 9: Zastoupení odpovědí na otázky z dotazníku v tab. 3 na škále *ano–ne*.



Obrázek 10: Zastoupení odpovědí na otázky z dotazníku v tab. 3 na škálách jiných než *ano-ne*.

při jejích různých opakováních stále stejnou odpověď (ať už správnou nebo špatnou). Tuto konzistenci jsme počítali následujícím vzorcem:

$$k = \frac{\sum_{p \in P} \frac{\max_{o \in O_p} \text{počet}(p, o)}{\text{opakování}(p)}}{|P|}$$

kde:

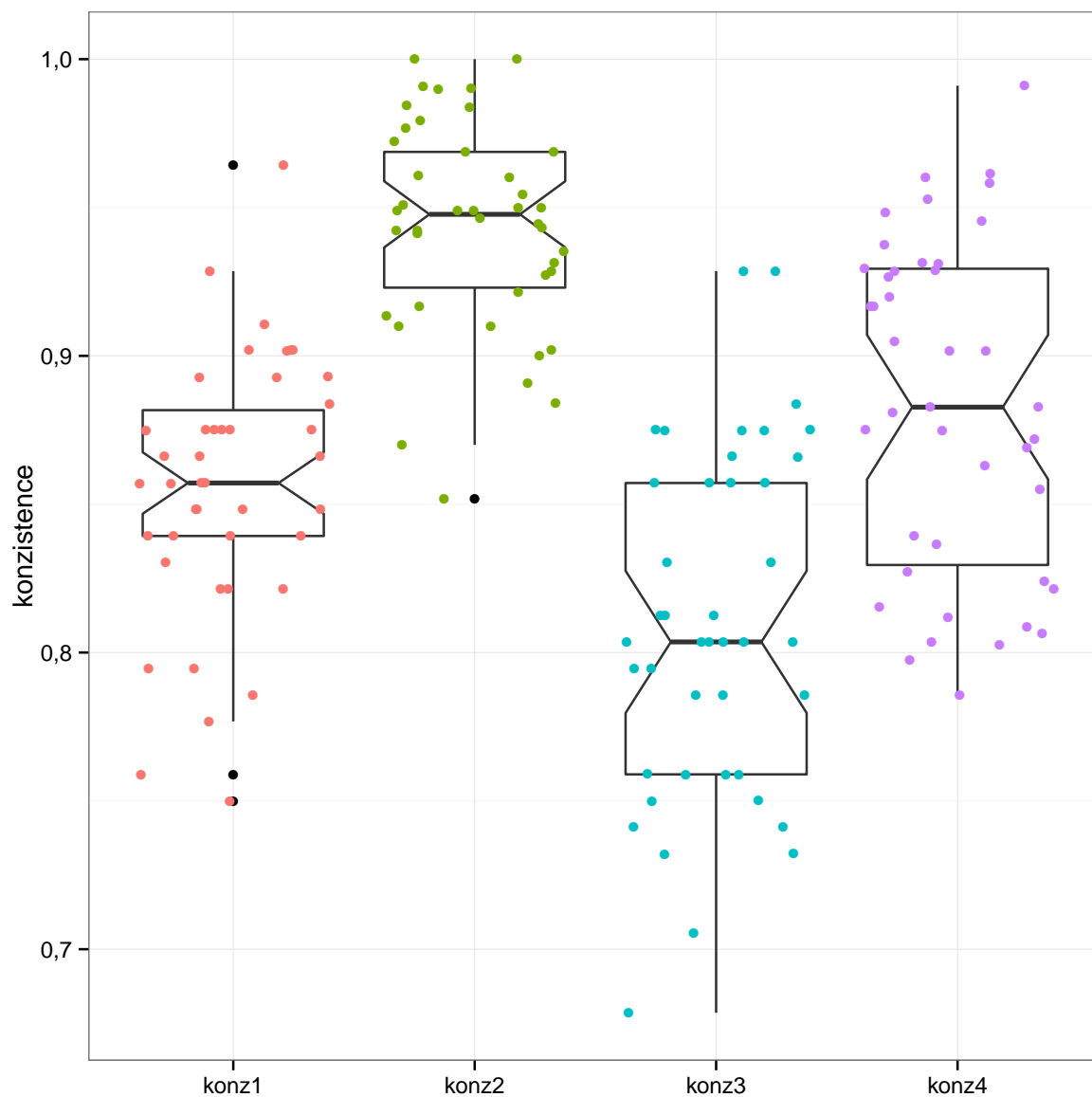
- P je množina testových položek a p jedna z nich
- O_p je množina možných odpovědí na položku p (v našem případě vždy $|O_p| = 2$) a o jedna z nich
- $\text{počet}(p, o)$ je počet případů, kolikrát respondent při různých opakováních položky p vybral odpověď o
- $\text{opakování}(p)$ je počet různých opakování položky p v rámci testu
- k je výsledná konzistence odpovědí subjektu

Rozpětí hodnot k je tedy $< 0,5; 1 >$, od čistě náhodných odpovědí, kde je očekávané rozdělení půl napůl ve prospěch obou nabízených variant, až po dokonale konzistentní odpovědi, kde subjekt na danou položku odpověděl pokaždé stejně, ať už dobře či špatně. Zároveň je toto dobrou příležitostí ověřit, zda požadavek, aby u každé otázky respondent ohodnotil subjektivní stupeň jistoty u odpovědi, splnil svůj účel. Pokud totiž do konzistence započítáme pouze otázky, u nichž respondenti uvedli, že se nerozhodovali náhodně, měla by celkově stoupnout.

Ukazuje se, že tomu tak skutečně je, jak shrnuje obr. 11. Pro pár vlevo (*konz1* a *konz2*) záleželo na pořadí nabízených odpovědí pro určení identity položky (AXB a BXA tedy byly dvě různé položky a každá z nich byla v rámci testu opakována dvakrát); u páru vpravo (*konz3* a *konz4*) na pořadí nezáleželo (AXB i BXA spadalo pod stejnou položku, dohromady opakovanou čtyřikrát). Levé členy párů (*konz1* a *konz3*) pak subjektivní jistotu nezohledňují, zatímco pravé členy (*konz2* a *konz4*) ano. Vidíme, že konzistence v obou párech měření vlivem zohlednění jistoty výrazně stoupne a rozdíl je signifikantní (vruby na krabicích se nepřekrývají).

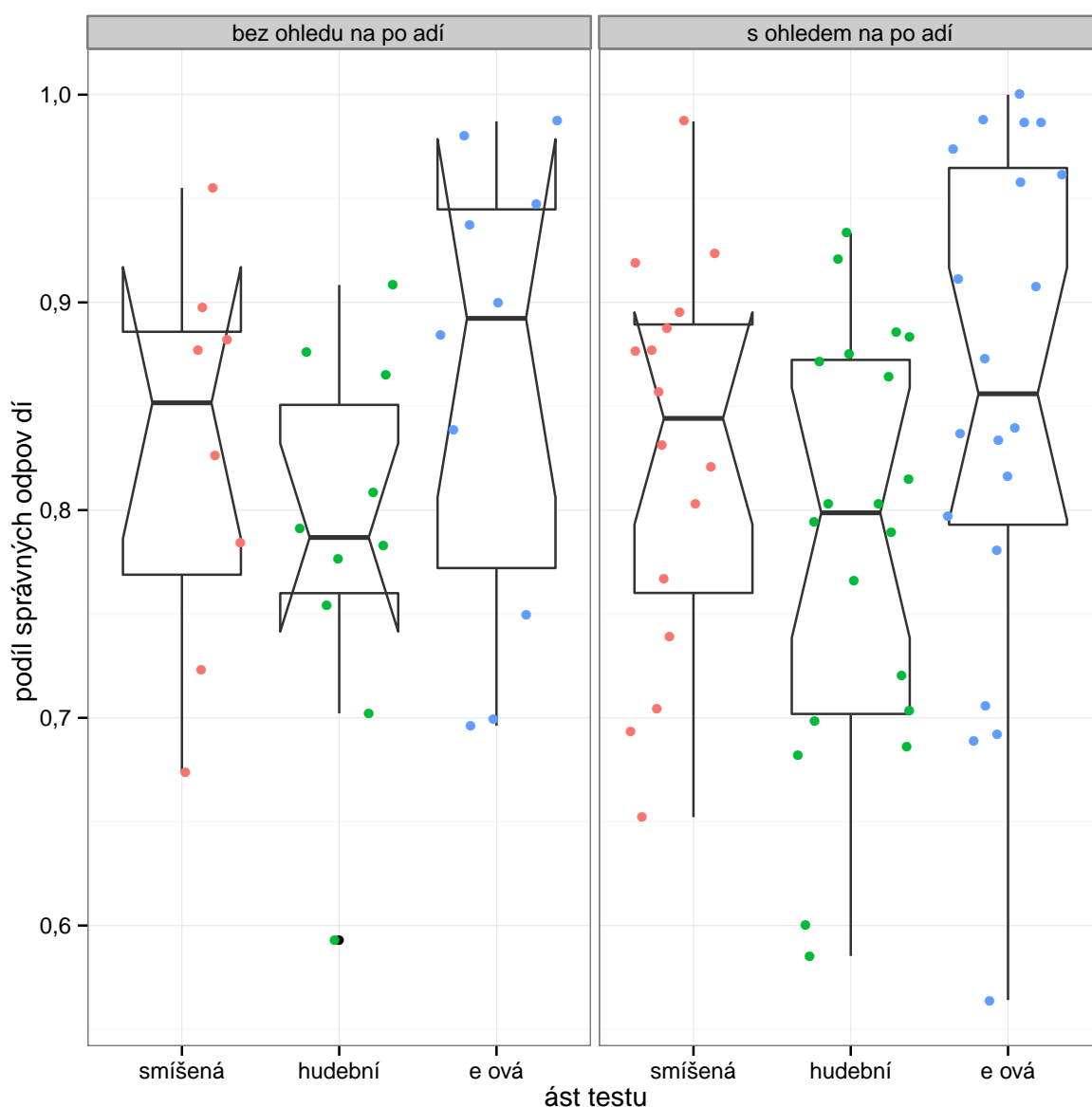
Z výše uvedeného plyne, že při vyhodnocování celkového skóre úspěšnosti testu podle vzorce popsaného v odd. 4.5 má skutečně smysl brát ohled na subjektivní stupeň jistoty uvedený respondentem spolu odpovědí, aby se odfiltroval šum způsobený náhodnými odpověďmi.

Obr. 12 pro zajímavost shrnuje, jak velký podíl (nenáhodných) odpovědí ve třech částech testu byl správně, tj. jak byly jednotlivé položky pro respondenty těžké či lehké. V levé půlce grafu je položka definována bez ohledu na pořadí nabízených odpovědí (AXB i BXA jsou tedy stejná položka), v pravé pak s ohledem na ně (AXB a BXA jsou považovány za dvě různé položky). Vidíme, že se dle záměru podařilo pokrýt širokou škálu obtížností.



Obrázek 11: Hodnoty konzistence odpovědí jednotlivých subjektů. Detaily viz text. Důležité je, že při omezení výpočtu konzistence na položky, u nichž respondenti uvedli, že neodpovídali náhodně (*konz2* a *konz4*) konzistence celkově stoupá oproti výpočtům založeným na všech položkách (*konz1* a *konz2*).

Řečová část se zdá obecně o něco lehčí a hudební naopak těžší, ale vzhledem k velké variabilitě hodnot ve všech třech částech se vruby překrývají a trend tedy není významný. U drtivé většiny položek je přitom podíl správných odpovědí statisticky významně vyšší, než by bylo očekáváno při náhodném vyplňování (= 0,5, při dvou nabízených možnostech)⁴⁸: všechny, kromě jedné hudební v přístupu bez ohledu na pořadí a dvou hudebních a jedné řečové v přístupu se zohledněním pořadí. Graf mimo jiné i nepřímo potvrzuje validitu smíšených stimulů (tj. to, že vztah mezi větou a hudebním analogem je skutečně motivovaný), neboť rozložení výsledků se ve smíšené části signifikantně neliší od ostatních dvou.



Obrázek 12: Podíl správných odpovědí na jednotlivé položky ve třech částech testu, počítáno jen z těch odpovědí, které respondenti označili za nenáhodné. Detaily viz text.

⁴⁸Testováno jednostranným binomickým testem s H_0 : „podíl správných odpovědí je roven 0,5“.

5.3 Korelační měření

Dostáváme se k hlavní části výsledků – korelacím, které mají ukázat, zda jsou řeč a hudba v lidské percepci do jisté míry provázané, a citlivost na jemné temporální či melodické variace v jedné doméně implikuje obdobnou schopnost i v doméně opačné. Na obr. 13, 14 a 15⁴⁹ jsou vynesena skóre respondentů ($N = 42$) v jednotlivých částech testu v závislosti na sobě navzájem, proložena regresní přímkou včetně 95% intervalu spolehlivosti.

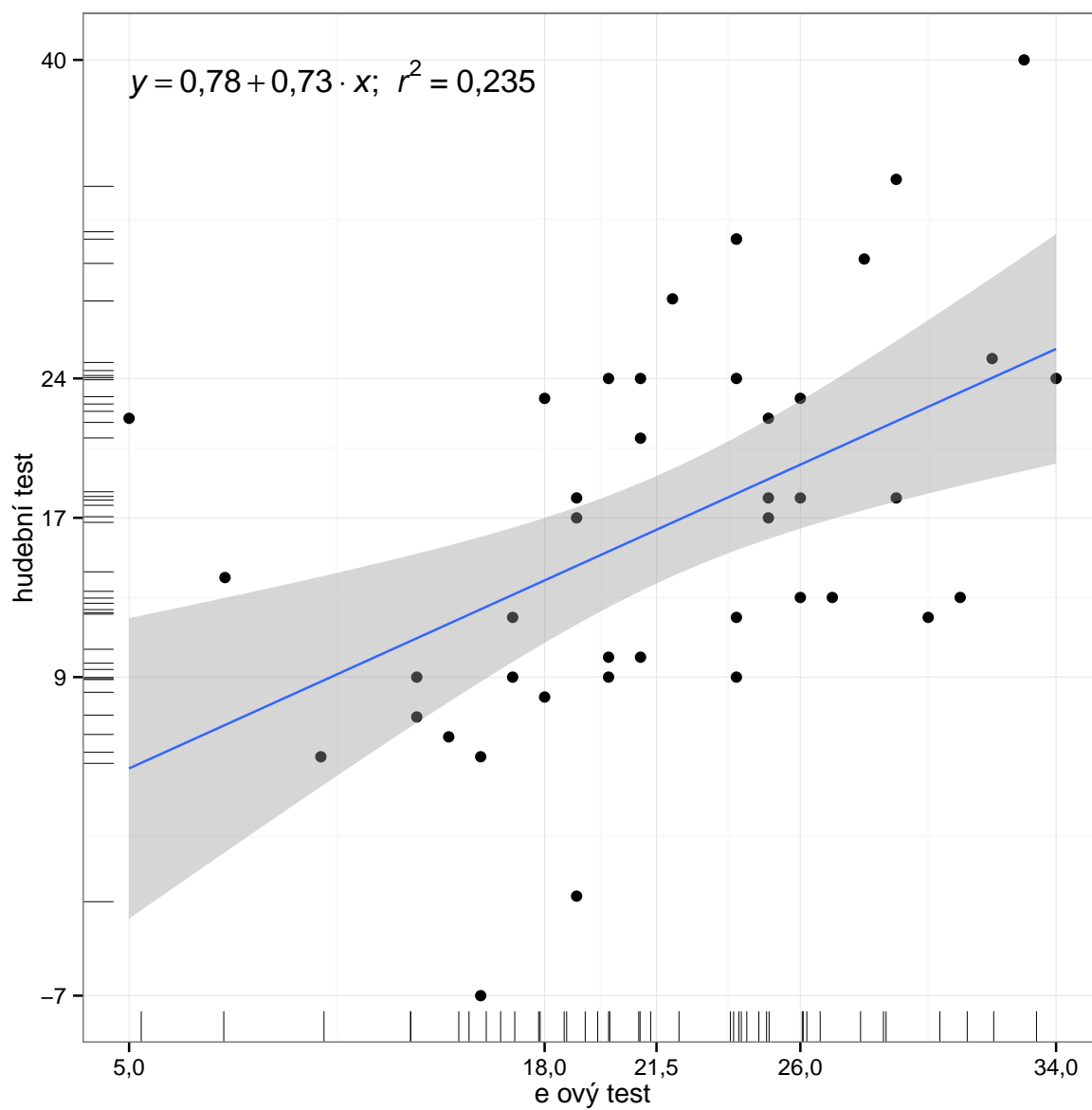
Ve všech třech grafech je patrné, že většina bodů sleduje očekávanou pozitivní korelaci mezi oběma veličinami; nejvíc výjimek je ve vztahu mezi řečovým a smíšeným skóre, což při stávající velikosti vzorku vede k intervalu spolehlivosti, který zahrnuje i horizontální sklon regresní přímky (tj. nulovou korelaci), a není tedy průkazný. Naopak nejsilnější a nejjasnější provázání je vidět mezi hudebním a smíšeným skóre (obr. 15). V grafech, které zahrnují skóre řečového testu (obr. 13 a 14) přitahují pozornost dvě odlehlé hodnoty s nápadně nízkým řečovým skóre a nečekaně vysokou úspěšností hudební či smíšenou. Zvláště v obr. 13 tyto body zamlžují jinak relativně přesvědčivé provázání obou veličin a snižují sklon regresní přímky oproti trendu vyplývajícímu ze zbytku datových bodů; taktéž snižují r^2 , tj. přiléhavost modelu.

Jeden z těchto dvou respondentů (ten s nižším řečovým, ale naopak vyšším hudebním a smíšeným skóre) dokonce přímo během testu (v pauze po první, řečové části) experimentátorovi sdělil, že se obává, že v několika otázkách omylem klikl naopak (tj. místo označení zvuku, který byl stejný jako referenční, vědomě označil ten, který se podle něj lišil). Vzhledem k tomu, že nejasnosti se zadáním měli i další respondenti, je možné, že výsledky v řečové části, která byla zařazena jako první, částečně utrpěly tím, že si řešitelé nejprve museli zvyknout na testové paradigma; u dvou jmenovaných bodů se pak tento efekt mohl projevit zvlášť markantně. Tato domněnka není dostatečně silným důvodem k úplnému vyřazení obou hodnot, ale u některých výpočtů níže pro srovnání uvedeme i hodnoty spočítané na základě podmnožiny měření, která by vznikla vypuštěním těchto respondentů z vzorku ($N = 40$).

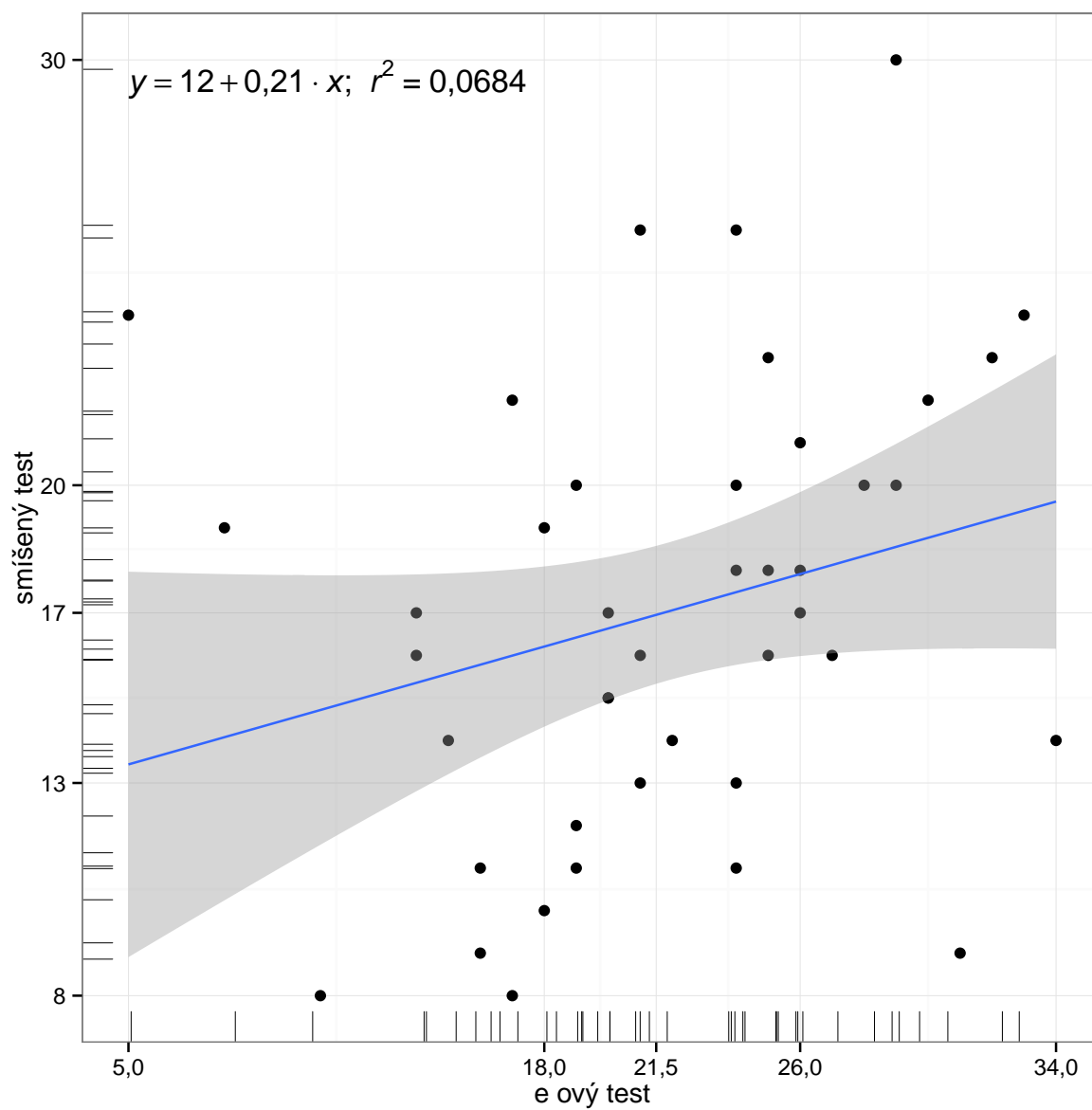
Vztah mezi řečovým a smíšeným skóre na obr. 14 je ovlivněn nejen odlehlými hodnotami respondentů, kteří si možná spletli zadání v první části, ale i nízkými hodnotami smíšeného skóre některých subjektů, kteří přitom byli v řečové části vysoce úspěšní (část grafu vpravo dole). Ukazuje to, že vztahu mezi větami a jejich hudebními analogy byly potřeba pro správné vyřešení úloh nejprve „přijít na kloub“, což ne všichni dokázali.

Jak je vidno z obr. 15, tato schopnost „přijít na kloub“ vztahu mezi větou a analogem mnohem lépe koreluje s úspěšností v hudebním testu. Jinými slovy, když už někdo uspěl

⁴⁹V těchto grafech kolmé „zářezy“ podél os naznačují rozložení pozorovaných hodnot v rámci dané proměnné; čísla na osách pak představují jejich tzv. pětičíselné shrnutí (minimum, 1. kvartil, medián, 3. kvartil, maximum). Graf též vždy obsahuje rovnici regresní přímky a r^2 , které kvantifikuje podíl rozptylu v datech vysvětlený modelem.

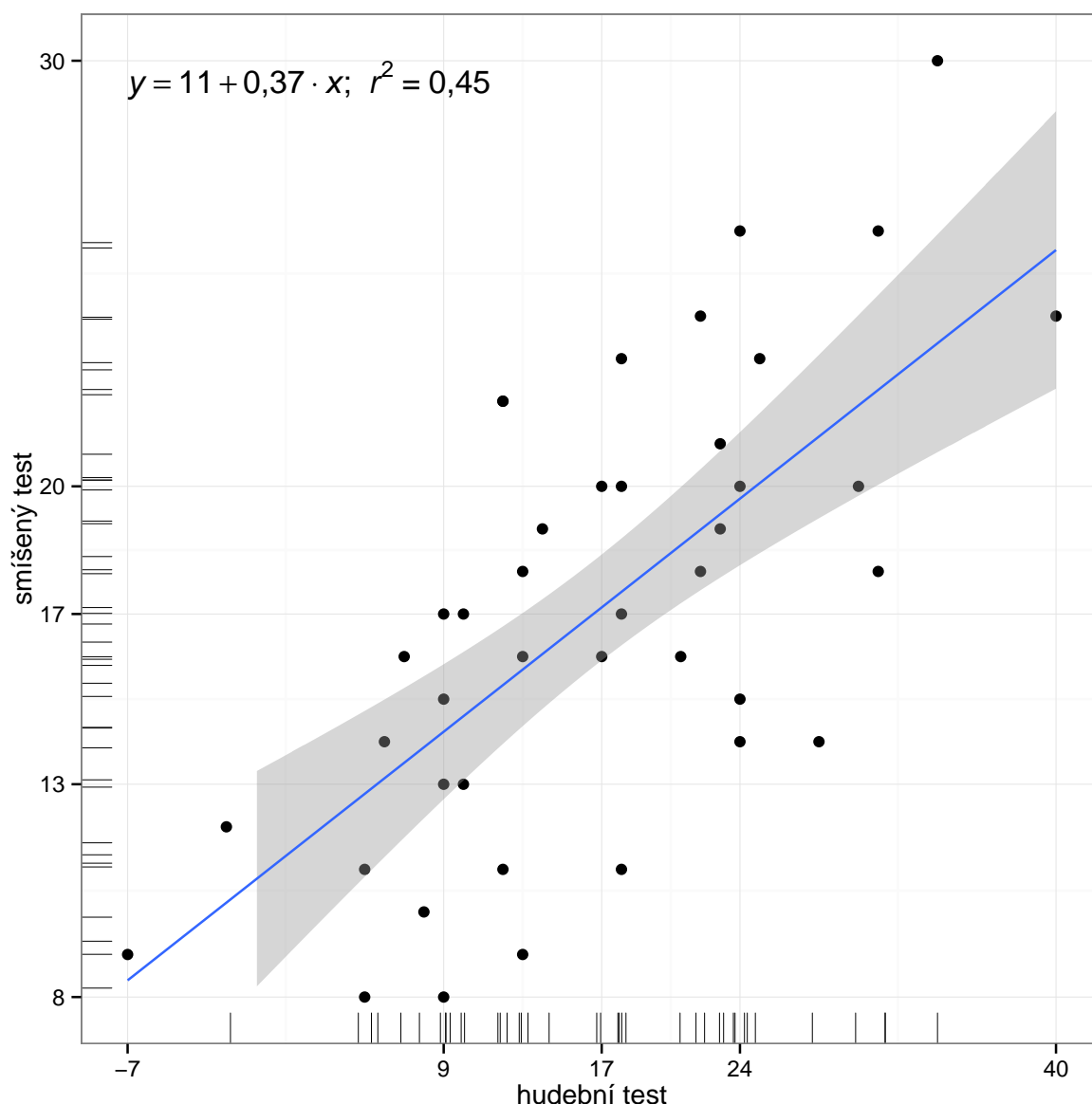


Obrázek 13: Grafické znázornění korelace mezi skóre jednotlivých respondentů v řečové a hudební části testu, včetně regresní přímky s 95% intervalem spolehlivosti.



Obrázek 14: Grafické znázornění korelace mezi skóre jednotlivých respondentů v řečové a smíšené části testu, včetně regresní přímky s 95% intervalem spolehlivosti.

v hudební části, bylo hodně pravděpodobné, že se mu v části smíšené podaří rozluštit paralely potřebné k stanovení správných odpovědí.



Obrázek 15: Grafické znázornění korelace mezi skóre jednotlivých respondentů v hudební a smíšené části testu, včetně regresní přímky s 95% intervalem spolehlivosti.

Co se týče číselné kvantifikace korelace, v tab. 4 uvádíme pro všechny tři vztahy naznačené grafech tři různé korelační koeficienty, včetně jejich statistické významnosti.⁵⁰ První z nich, Pearsonovo r , detekuje lineární závislost mezi proměnnými, tj. míru, v jaké dvě

⁵⁰Symbody pro statistickou významnost na různých hladinách α použité v tomto textu jsou následující: . pro hladinu $\alpha = 0,1$ (okrajová významnost, viz dále), * pro 0,05, ** pro 0,01 a *** pro 0,001. „Okrajová významnost“ je poněkud kontroverzní pojem, je tedy potřeba si uvědomit, co přesně znamená. Nejde totiž o výsledek, který se skoro blíží významnosti, nýbrž o výsledek, který bychom určili za významný, kdybychom obecně byli při testování hypotéz ochotni připustit větší procento chyb prvního druhu (nesprávně zamítnutá H_0).

proměnné rostou či klesají rovnoměrně spolu (pozitivní korelace) či proti sobě (negativní korelace). Při menším počtu datových bodů ovšem může pár odlehlých hodnot linearitu vztahu silně narušit, jak jsme si ukázali na obr. 13 a 14. Proto uvádíme i další dva koeficienty, Kendalovo τ a Spearmanovo ρ , které jsou založené na korelaci pořadí, tudíž nevyžadují lineární vztah mezi proměnnými⁵¹ a jsou méně citlivé na odlehlé hodnoty.

vztah	koef.	hodnota	95% int. spoleh.	významnost	síla
hudba ~ řeč	r	0,48	0,21–0,69	**	střední
	τ	0,41	NA	***	střední
	ρ	0,56	NA	***	střední
smíšené ~ řeč	r	0,26	-0,05–0,52	.	nízká
	τ	0,26	NA	*	nízká
	ρ	0,33	NA	*	nízká
smíšené ~ hudba	r	0,67	0,46–0,81	***	střední
	τ	0,49	NA	***	střední
	ρ	0,66	NA	***	střední

Tabulka 4: Korelace mezi skóre respondentů v různých částech testu (sloupec *vztah*), na základě tří různých korelačních koeficientů (sloupec *koef.*: Pearsonovo r , Kendalovo τ a Spearmanovo ρ). Sloupec *významnost* uvádí, zda se hodnota korelačního koeficientu statisticky významně liší od 0. Sloupec *síla* obsahuje slovní hodnocení síly efektu korelace podle Volína (2007, s. 190).

Tab. 5 pak pro zajímavost uvádí přepočítané hodnoty stejných korelačních koeficientů, pokud ze sady měření odstraníme dvě výše zmíněné odlehlé hodnoty s nízkým skóre v řečové části. Vidíme značný nárůst zejména Pearsonova r , neb odstraněním odlehlých hodnot stoupá linearita vztahu; citelný je dopad především na vztah mezi řečovým a smíšeným skóre.

vztah	koef.	hodnota	95% int. spoleh.	významnost	síla
hudba ~ řeč	r	0,60	0,36–0,77	***	střední

⁵¹Např. je-li vztah mezi proměnnými kvadratický, stále se jedná o monotónní růst (za předpokladu, že je koeficient u kvadratického členu pozitivní) a $\tau = \rho = 1$ (dokonalá pozitivní korelace). Naopak $r < 1$.

vztah	koef.	hodnota	95% int. spoleh.	významnost	síla
	τ	0,47	NA	***	střední
	ρ	0,64	NA	***	střední
smíšené ~ řeč	r	0,45	0,16–0,67	**	střední
	τ	0,35	NA	**	nízká
	ρ	0,46	NA	**	střední

Tabulka 5: Hodnoty z tab. 4 přepočítané pro vzorek, z něž byly vypuštěny dvě odlehlé hodnoty v řečových skóre.

5.4 Lineární modely: hledání faktorů úspěšnosti

Aby bylo možné blíže a komplexněji prozkoumat, které faktory mají na úspěch v jednotlivých částech testu, byl pro každé ze tří skóre vytvořen lineární model, který predikuje⁵² jeho hodnotu současně na základě následujících čtyř proměnných⁵³:

- skóre ve zbývajících částech testu (tj. např. hudební skóre se snažíme predikovat na základě řečového a smíšeného)
- kategorie *hudebnost* popsané v odd. 4.5, stanovené na základě odpovědí v dotazníku v tab. 3; pět úrovní a–e (v pořadí stoupající „hudebnosti“)
- míry, do níž je VŠ absolvovaná respondentem zaměřená na vědomou reflexi jazyka, se zvláštním zřetelem k jeho zvukové stránce (taktéž podle dotazníku); pět úrovní (ve stoupajícím pořadí):

1. VŠ není nijak jazykově zaměřena
2. mezinárodní vztahy⁵⁴

⁵²Uvedme hned zkráj, že slova „predikovat“ či „prediktor“ používáme čistě jako technický termín: záměrem není vytvořit model, který by na základě nových dat skóre skutečně predikoval, ale popsat data stávající. Lineární model je pro nás pouze matematickým mechanismem, který umožňuje integrovat a porovnávat vztahy a sílu provázání jedné proměnné s vícero dalšími. Nejde nám tedy o kauzalitu, nýbrž znovu o korelaci; pro zkoumání kauzality by bylo potřeba provést mnohem lépe kontrolovaný experiment.

⁵³Interakce jsme do modelů žádné nezahrnuli, zařazení příliš velkého počtu prediktorů při stále relativně malém počtu datových bodů ($N = 42$) by výsledný model mohlo znehodnotit. Navíc je málo pravděpodobné, že by se při této velikosti vzorku a omezeného počtu kombinací úrovní prediktorů, které nabízí, mohly nějak výrazněji projevit.

⁵⁴Studium mezinárodních vztahů implikuje intenzivní kontakt s cizími jazyky a vede tedy k vytvoření referenčního rámce, který může být podkladem pro zostřené vnímání obecných aspektů lidské řeči jako takové.

3. filologie
4. obecná lingvistika
5. fonetika

První dva použité prediktory jsou numerické a charakterizují percepční citlivost v hudební či řečové doméně projevenou v praxi přímo v rámci testu. Druhé dva jsou ordinální a souvisí naopak s respondentovou předpokládanou teoretickou obeznámeností s oběma doménami, stanovenou na základě jím deklarované předchozí praxe s nimi. Než ovšem přikročíme k rozboru lineárních modelů, podívejme se nejprve detailněji na vztah úspěšnosti v jednotlivých částech testu s oběma ordinálními prediktory (vzájemný vztah dosaženého skóre mezi částmi testu jsme již popsali v předchozí části o korelacích).

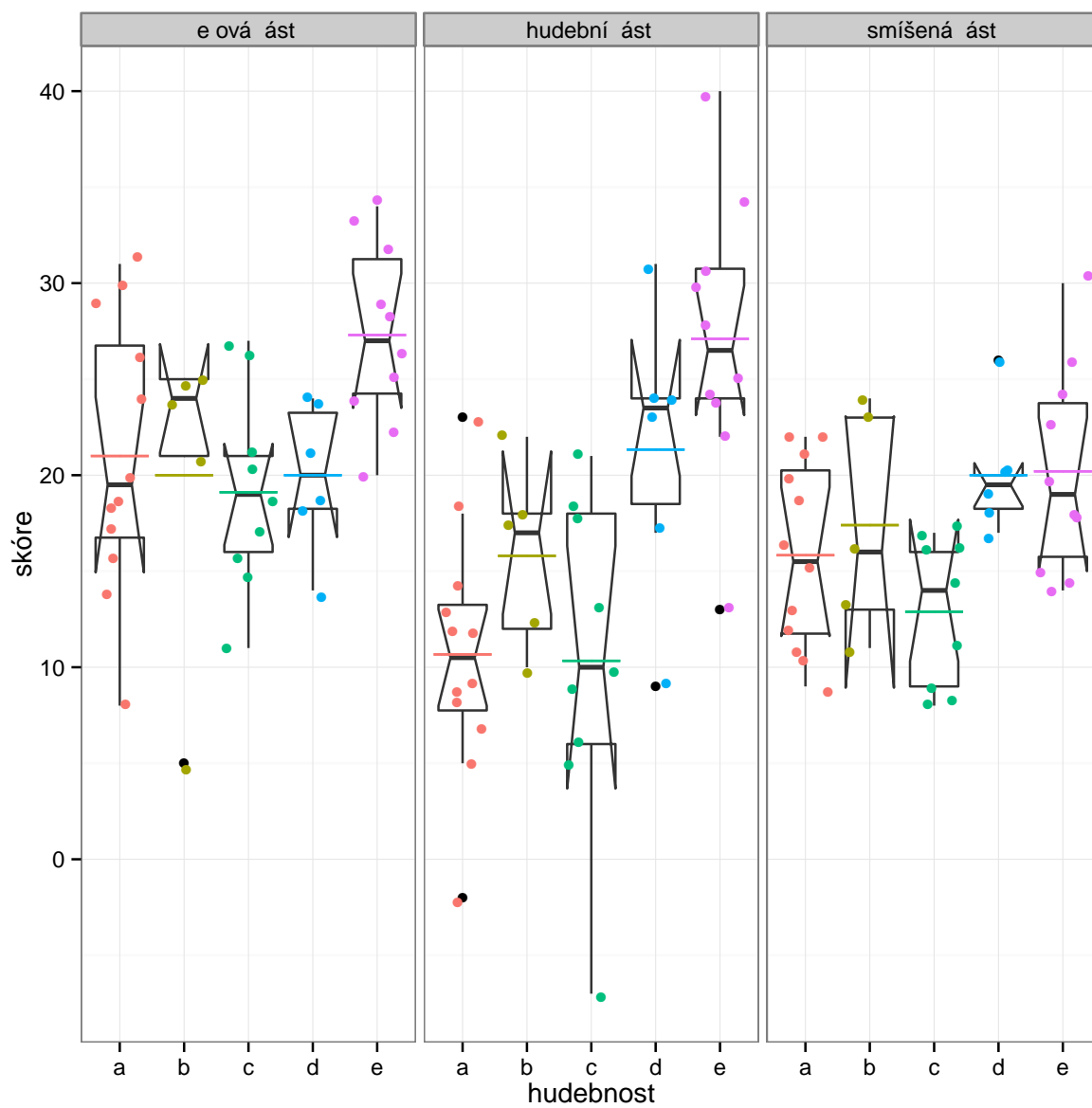
Vliv *hudebnosti* na řečová, hudební a smíšená skóre shrnuje obr. 16. Jistá souvislost mezi hudebností a skóre je patrná zejména v hudební části, ve smíšené části se naopak a možná překvapivě projevuje daleko méně, přestože korelace mezi hudebním a smíšeným skóre byla jasně nejvyšší. To naznačuje, že teoretické hudební zázemí nebylo nutně klíčem k dobrému výsledku v testu. Všech pět kategorií a–e se nicméně poměrně často v pozorovaných hodnotách překrývá, což neumožňuje spolehlivou vizuální interpretaci významnosti rozdílů mezi nimi. Přikročili jsme tudíž u každé části k jednofaktorové analýze rozptylu (ANOVA) skóre v závislosti na *hudebnosti*. Nezajímal nás přitom hlavní efekt, ale to, které kontrasty mezi kategoriemi jsou statisticky významné, tj. které kategorie jsou jasně odlišené a které spolu naopak splývají, což jsme stanovili pomocí *post-hoc* testu Tukeyho HSD.

V řečové části se zdá, že skupina s nejvyšší hudebností má tendenci mít vyšší skóre, ale signifikantní je pouze kontrast $e \times c$, což nehovoří o příliš přesvědčivém propojení (u ordinální proměnné bychom očekávali, že kontrast $e \times c$ bude implikovat i kontrasty $e \times a$, $e \times b$).⁵⁵ Podobně neprůkazná je situace u smíšených stimulů (znovu jen rozdíl $e \times c$). Naopak v hudbě čísla potvrzují předchozí vizuální dojem: významné jsou kontrasty $e \times a$, $e \times b$, $e \times c$; $d \times a$, $d \times c$. Kontrast $d \times b$ sice chybí, ale celkově se zdá, že v hudební části máme dvě skupiny úspěšnosti podle *hudebnosti*: $\{a, b, c\}$ a $\{d, e\}$.⁵⁶

Stejnou metodou jsme stanovili významnost kontrastů úrovní prediktoru i u vztahu mezi *VŠ* a skóre, zobrazenému na obr. 17. Zde je třeba podotknout, že ve dvou kategoriích (mezinárodní vztahy a lingvistika) bylo respondentů velmi poskrovnu (3, respektive 2), a kontrasty s nimi proto v podstatě neměly šanci vyjít významné, neboť byl vzorek příliš

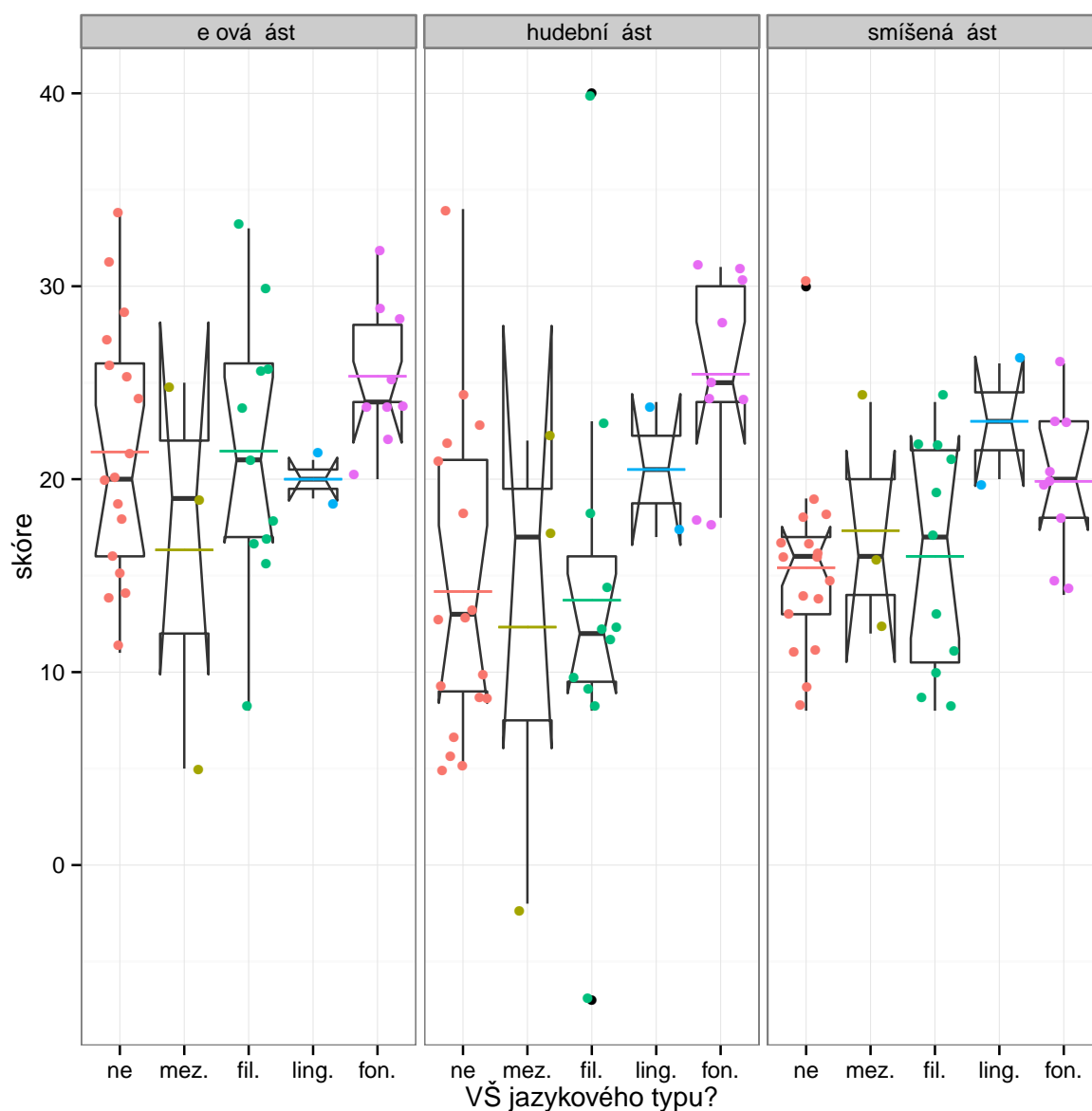
⁵⁵Možná by se vztah projevil při vyšším počtu datových bodů, Tukeyho HSD totiž snižuje cílovou hladinu α u párových porovnání tak, aby jejich výsledná kombinovaná hladina α splňovala zvolené požadavky na významnost, čímž činí dosažení signifikantního výsledku těžším.

⁵⁶Z hlediska některých lineárních modelů by tedy bylo možné některé kategorie prediktoru *hudebnost* sloučit (podobně – byť v menší míře – viz níže pro prediktor *VŠ*). V praxi se ovšem ukázalo, že takový postup nevede v ohledech pro nás zajímavých k odlišným výsledkům.



Obrázek 16: Skóre v jednotlivých částech testu podle stoupající *hudebnosti* subjektu (a–e). Barevné horizontální čáry naznačují aritmetický průměr v rámci kategorie.

malý. Jediné dva významné kontrasty se vyskytly v hudební části: fon. \times ne a fon. \times fil. To sice zdánlivě hovoří pro možnost vydělení fonetiků, případně fonetiků a lingvistů; absenci kontrastu fon. \times mez., kterou bychom znovu očekávali vzhledem k ordinalitě prediktoru, bychom mohli zanedbat, neb je očekávatelná kvůli výše zmíněnému malému počtu datových bodů v mez. Nicméně se ukazuje, že valná většina fonetiků má zároveň vysokou *hudebnost* ($5 \times e$, $2 \times d$ proti $1 \times a$, $1 \times b$), takže pod povrchem tohoto kontrastu bude zřejmě ve skutečnosti již výše popsáný vztah mezi *hudebností* a hudebním skóre. Úhrnem tedy předběžně můžeme odhadnout, že vliv předchozího lingvisticky či přímo foneticky zaměřeného vzdělání byl zanedbatelný, a to dokonce i v řečové části.



Obrázek 17: Skóre v jednotlivých částech testu podle stoupající hodnoty prediktoru VŠ u subjektu (ne–fon., blíže viz text). Barevné horizontální čáry naznačují aritmetický průměr v rámci kategorie.

Přesuňme se k samotným lineárním modelům. Předně, aby byl konkrétní vypočítaný model validní a bylo možné jej interpretovat, musí splňovat jistá kritéria. Ta byla částečně ověřena funkcí `glvma` ze stejnojmenného balíčku ve statistickém prostředí R (R Core Team 2014),⁵⁷ v němž byly modely počítány, částečně vizuální inspekci dat v grafu a částečně vyplývaly přímo z povahy dat. Šlo zejména o:

- vlastnosti reziduálů (tj. odchylek reálných bodů od hodnot predikovaných modelem): homoskedasticita (reziduály by v závislosti na predikovaných hodnotách měly kolísat čistě náhodně) a normalita rozdělení
- minimum odlehklých bodů („příliš vlivných“ hodnot)
- nezávislost pozorování: zde je vše v pořádku již z principu, každý datový bod pochází od jiného respondenta
- absence korelace mezi prediktory: zde naopak víme, že skóre jsou mezi sebou slabě až středně silně korelovaná, a navzdory tomu jako prediktory vždy používáme dvě z nich; musíme se s tím ovšem smířit a spolehnout se na to, že korelace není až tak výrazná, neboť standardní řešení této situace spočívá v převedení takových prediktorů na nekorelované hlavní komponenty pomocí metody PCA, přičemž umístění respondenta na škále těchto komponent by již ztratilo jasný vztah k výkonu v testu a zkomplikovalo interpretaci

Vyhodnocení lineárních modelů pak probíhalo tak, že z nich byly postupně odebírány prediktory, jejichž vyřazení signifikantně neovlivňovalo celkovou přesnost modelu. Tab. 6⁵⁸ shrnuje jejich výslednou podobu; pro zajímavost obsahuje i modely spočítané s vypuštěním dvou odlehklých hodnot popsaných v odd. 5.3.

model	vynech.?	F	df	celk. sig.	r^2	prediktory	sig. pred.
řeč. skóre	ne	12,30	1; 40	**	0,2160	<i>hud. skóre</i>	**
řeč. skóre	ano	7,11	5; 34	***	0,4392	<i>hud. skóre</i>	***
						<i>hudebnost</i>	.
hud. skóre	ne	13,49	6; 35	***	0,6464	<i>smíš. skóre</i>	***
						<i>hudebnost</i>	**
						<i>řeč. skóre</i>	.
hud. skóre	ano	14,82	6; 33	***	0,6801	<i>hudebnost</i>	**

⁵⁷Odhalila pouze nevyhovující špičatost rozdělení u modelu řečového skóre bez vypuštění dvou odlehklých hodnot popsaných v odd. 5.3. Vzhledem k tomu, že při odstranění těchto hodnot ze vzorku tento problém zmizel a model vyšel podobný (viz tab. 6), jsme se rozhodli na něj nebrat ohled.

⁵⁸Symboly pro různé úrovně statistické významnosti jsou popsány v poznámce pod čarou na s. 74.

model	vynech.?	F	df	celk. sig.	r^2	prediktory	sig. pred.
						<i>řeč. skóre</i>	**
						<i>smíš. skóre</i>	**
smíš. skóre	ne	32,78	1; 40	***	0,4367	<i>hud. skóre</i>	***
smíš. skóre	ano	31,78	1; 38	***	0,4411	<i>hud. skóre</i>	***

Tabulka 6: Tabulka charakteristik výsledných lineárních modelů. Sloupce popisují (zleva doprava): modelovanou proměnnou; zda byly vynechány některé odlehle hodnoty; testové kritérium F pro celý model; stupně volnosti; celkovou významnost modelu; upravené r^2 („přiléhavost“ modelu k původním hodnotám); ponechané prediktory (z původních čtyř, viz seznam na začátku tohoto odd.); významnost daného prediktoru.

Na první pohled zaujme, že se ani v jednom z případů mezi významné prediktory nedostal faktor *VŠ*. Taktéž je patrné, že vyřazení dvou odlehlých hodnot jasně zvyšuje užitečnost řečového skóre jako prediktoru: v modelu hudebního skóre se tímto krokem posouvá z okrajové významnosti na významnost plnou. Nicméně v modelu smíšeného skóre není signifikantním prediktorem ani tak, což jen potvrzuje to, co jsme říkali již v komentáři ke korelacím: jak ukazuje obr. 14, někteří jedinci navzdory vysokému řečovému skóre nebyly schopni „přijít na kloub“ paralelám mezi větami a jejich hudebními analogy. Naopak úspěch v hudební části tuto schopnost předznamenával relativně spolehlivě (viz obr. 15). Další postřehy poskytne následující kapitola.

6 Diskuse

6.1 Vliv kategorických proměnných *hudebnost* a *VŠ*

Obě kategorické proměnné se svým dopadem silně liší, lépe řečeno *hudebnost* hraje významnou roli pro úspěch v hudebním testu, kdežto jazykově zaměřená *VŠ* se nezdá být relevantní ani pro řečovou část. To je povzbudivá zpráva, neboť z toho můžeme dovodit, že se nám podařilo vytvořit řečové úlohy, k jejichž správnému řešení není nutné mít speciální jazykovědnou průpravu, a testují schopnosti, které má každý uživatel jazyka běžně k dispozici, byť v různé míře.

Při pohledu na první panel obr. 17 stojí za pozornost, že ač fonetici nebyli v řečovém testu jediní dobří (někteří respondenti z první kategorie mají srovnatelné výsledky, jeden z nich dokonce nejvyšší řečové skóre vůbec), nejsou mezi nimi vyloženě špatné výsledky. V kategorii fon. je zdaleka nejmenší rozptyl hodnot, pokud nepočítáme mizivě zastoupenou kategorii ling. Tento fakt nejspíš reflektuje buď vliv fonetické průpravy, nebo to, že do studia fonetiky se nepustí někdo, kdo by sám o sobě nedisponoval jistou předchozí citlivostí na zvukovou stránku řeči (případně kombinaci obojího).

Významná role *hudebnosti* pro úspěch v hudebním testu je do jisté míry překvapivá. Slevc a Miyake v již zmíněném výzkumu (2006) také mimo jiné žádali subjekty, aby sami sebe ohodnotili stran hudební dovednosti, ale významným prediktorem úspěchu v různých jazykových úkolech bylo pouze skóre dosažené v praktickém hudebním testu. Možným klíčem k tomuto rozporu je, že naše *hudebnost* je objektivnější měrou hudebních dovedností než ono „sebeohodnocení“. Slevc a Miyake neuvádějí o jeho povaze žádné detaily, ale nazývají jej subjektivním, možná se tedy jednalo dokonce o jednu jedinou otázku typu „Ohodnoťte svou hudební zdatnost na škále od 1 do 5.“ Odpověď na takovou otázku záleží na mnoha faktorech: jak si kdo škálu interpretuje, jaké má kdo sebevědomí apod. Naproti tomu naše *hudebnost* byla odvozená na základě otázek, které měly monitorovat míru hudební aktivity vyvíjené respondentem.

Zde se pochopitelně nabízí ošemetná otázka, zda je pak dobrý výkon v hudebním testu důsledkem vzdělání a praxe či přímo drilu spojeného s touto aktivitou, či zda hraje roli spíš to, že když se někdo hudbě intenzivně věnuje, pravděpodobně k ní má i nadání (dá se předpokládat, že málokdo tak činí přímo nedostatku hudebního talentu navzdory). Nejspíš půjde o kombinaci obojího, přičemž výhoda drilu se ovšem projeví pouze v úlohách, které přímo odpovídají oblasti, v níž drill probíhá, tj. v hudebním testu. Ve smíšeném či dokonce řečovém testu, kde přitom šlo o podobné druhy manipulací, již *hudebnost* významným prediktorem nebyla; naopak jím byl reálný výkon v hudebním testu, tedy prokázaná percepční citlivost bez ohledu na vzdělání.

6.2 Provázání řečové a hudební percepce

Zkraje znovu zdůrazněme, že zatímco významnost korelací mezi jednotlivými částmi testu je uspokojivá (příčemž u vztahu mezi řečí a smíšenými stimuly to platí spíše až při vypuštění dvou odlehlých hodnot), síla efektu už tak výrazná není: pohybuje se většinou v oblasti středně silné korelace. Nejslabší vztah je mezi smíšeným a řečovým testem za použití celého vzorku, zde je korelace jen nízká. Hodnoty neoptimističtějších odhadů se pohybují od 0,33 do 0,67 (viz tab. 4 a 5) a odpovídají většinou Spearmanovu ρ . To se počítá stejně jako Pearsonovo r , ovšem místo původních hodnot do vzorce dosadíme jejich pořadí. Tím se vyhladí jisté nepravidelnosti v růstu obou proměnných a zvýší se tak linearita vztahu, takže dává smysl, že je ρ až na jeden případ (korelace smíšené ~ hudba) vyšší než r počítané přímo z původních hodnot.

Naopak Kendallovo τ převážně zaostává. Jeho výpočet je založen čistě na porovnávání pořadí dvojic bodů: pokud je $x_i < x_j$ a zároveň $y_i < y_j$ (či je vztah v obou případech zase opačný), pak jde o dvojici souhlasnou a koeficient se jejím vlivem zvyšuje; jinak se snižuje. Z toho můžeme vyvodit, že v rámci pořadí subjektů dochází relativně často k prohození, která mají na τ negativní vliv; graficky se ostatně promítají i do poměrně širokých pásem, v nichž body na obr. 13, 14 a 15 leží. Tento rozptyl má vliv pochopitelně i na ρ a r (aby $r = 1$, musely by body ležet na jedné přímce), ovšem viditelně menší. Replikace experimentu s větším počtem subjektů by tedy pravděpodobně více vyplnila tento pás, ale nezúžila by ho. K tomu by nejspíš bylo nutné ještě vylepšit či zpřesnit experimentální paradigma; některé návrhy podáme v odd. 6.3.

Skóre ve smíšené části mnohem lépe koreluje se skórem v hudební než v řečové; i z hlediska lineárních modelů je při jeho predikci jediným relevantním faktorem právě hudební skóre. Trend pozitivní korelace je u vztahu smíšené ~ řeč sice v náznacích patrný, nicméně je zastíněný odlehlými hodnotami (kromě oněch dvou, které jsme pro srovnání v některých výpočtech vyřadili, i dalšími, jež nelze ignorovat ani takto tentativně). Abychom mohli vyhodnotit, zda v tomto ohledu skutečně mezi hudebními a řečovými percepčními vlohami existuje korelací naznačená asymetrie, která by zpochybňovala naši hypotézu o jejich propojení, museli bychom mít údaje i o opačné úloze.

Je ovšem otázkou, zda je takovou úlohu vůbec možné koncipovat. Podle nás se nejedná o prosté prohození rolí vět a jejich hudebních analogů (tj. použít analogy jako referenční zvuky a dát na výběr ze dvou vět, z nichž jedna bude manipulovaná), které jsme zde v rámci pilotáže kvůli délce testu a neintuitivnosti úlohy nakonec nezařadili. Adaptaci a luštění by v takovém případě totiž nadále vyžadoval hudební zvuk, ne jeho deklarovaný řečový „analog“. Je vůbec možné začít v tomto ohledu skutečně od melodie a vytvořit podle ní od základu řečový analog? Pak je zde také problém, že řečí vládne drtivá většina lidí velmi dobře, neb je základním požadavkem sociálního začlenění. U hudby tomu tak

není a rozdíly mezi jedinci jsou markantnější. Provázaná percepční citlivost neznamená, že lze postavit rovnítko mezi to, co se skrze ni jedinci v obou doménách učí: řečové stimuly budou vždy povědomější a přístupnější větší části respondentů, jednoduše protože mají s touto doménou mnohem více zkušeností.⁵⁹ Není tedy překvapivé, že ve smíšené části, která nejprve vyžaduje rozkódovat neznámý systém hudebních paralel, lépe uspěli zejména ti, kteří prokázali zdatnost i v čistě hudební oblasti.

S tím souvisí i fakt, že stimuly použité v našem výzkumu byly již poměrně složité (tři mluvni, resp. hudební takty), vyžadovali kromě základní citlivosti i jisté syntaktické zpracování. Znovu se tím vracíme k bazálním a kapitálním procesům (viz shrnutí na začátku kap. 3): správné a účinné zpracování zvukového vstupu vyžaduje jak spolehlivou analýzu dat (základní percepční citlivost zprostředkovanou bazálními drahami), tak efektivní schopnost predikce, aby byl mozek schopen odhadnout, jak načasovat fluktuaci soustředění tak, aby byl v toku dat schopen zacílit na potenciálně informačně relevantní místa (kapitální dráhy).⁶⁰ Je tedy pravděpodobné, že na úrovni kratších a syntakticky jednodušších jednotek by byly kladeny menší nároky na predikci a mohla by se více projevit základní citlivost. Tím by byli méně znevýhodněni respondenti, kteří mají sice „bystré ucho“, ale nijak je z hlediska hudby nepěstují.

Znevýhodnění je největší právě u smíšených stimulů, neboť vyžadují syntaktickou analýzu obou domén paralelně a navíc jejich porovnání, takže jsou na tom o mnoho lépe ti, kteří s tou méně obvyklou (hudbou) mají větší zkušenosti. Je toto dokladem pro Patelovu hypotézu o sdílených neurálních prostředcích pro syntaktickou integraci (SSIRH), či spíše jejím vyvrácením? Mohlo by se zdát, že tento fakt hovoří spíše proti ní: pokud je syntax obou domén propojená, neměl by se u smíšených stimulů promítnout i výkon v řečové části? Ne nutně, při zpracování syntaxe je potřeba operovat s mentálními reprezentacemi jednotek, přičemž o těchto reprezentacích Patel připouští, že jsou pochopitelně pro obě domény oddělené. U hudebně zdatnějších lidí se pak dá předpokládat větší mentální hudební „lexikon“ a tedy i snazší zpracování syntakticky složitějšího stimulu.

Připomeňme v tomto ohledu výzkum Fedorenkové et al. (2009), která zjistila, že hudebně-řečové stimuly (zpívané věty), které jsou z hlediska jazykové i hudební syntaxe inkongruentní, představují pro posluchače mnohem větší kognitivní zátěž, než by vyplývalo z pouhého součtu působení obou inkongruencí odděleně. Fedorenková z toho vyvodila, že zpracování řečové i hudební stránky je silně propojené a zablokování této dráhy⁶¹ jednou inkongruencí těsně následovanou další vede k jakémusi řetězovému zpomalení. Naše data

⁵⁹Nabízí se otázka, jak by v podobném testu tedy uspěl např. kmen Mbendjele zmíněný v odd. 1, pro nějž je hudba podle všeho srovnatelně důležitým společenským médiem, a u jehož příslušníků lze tedy očekávat podobně rozvinuté hudební dovednosti.

⁶⁰Opozici si lze také představit v intencích vztahu mezi fonetikou a fonologií.

⁶¹Nemusí jít nutně o jednu jedinou dráhu; může se jednat i o celou propojenou distribuovanou síť, v duchu pojetí Abramse et al. (2011) popsáno v odd. 2.2.1.

ze smíšeného testu by bylo možné v tomto světle interpretovat tak, že obeznámenost s hudbou (odrážející se v úspěšnosti v hudební části) vede k větší průchodnosti této dráhy pro zvuky hudebního charakteru, a tudíž i k možnosti lépe (bez zádrhelů) zpracovat také zvuky řečové, což se promítne do celkově vyššího počtu správně odhalených analogií.

6.3 Metoda a další výzkum

Během testování poskytli různí respondenti mnoho cenné zpětné vazby, která by mohla být hodnotným vodítkem pro toho, kdo by chtěl na výzkum navázat, či jej s jistými obměnami replikovat. Opakovaně se objevila např. připomínka, že měl respondent tendenci v položkách vybírat místo zvuku, který byl stejný jako referenční, naopak ten, který se lišil. Standardní paradigma AXB přitom převážně od subjektu žádá, aby se na úkol díval tak, že hledá podobnost, ne rozdíl. Je ovšem možné, že v rámci úloh použitých v tomto testu je hledat naopak rozdíl intuitivnější. Pokud standardní postup skutečně přidává kognitivní zátěž, která přímo nesouvisí se stimulem, a může tedy způsobit nechtěný šum v datech (viz potenciální odlehle hodnoty popsane v odd. 5.3), nevidíme důvod, proč úlohu neformulovat opačně.

Poměrně velká část respondentů uvedla, že pro ně bylo těžké porovnat referenční stimul s oběma neznámými najednou; většinou si tedy pustili každou položku dvakrát a při každém z obou poslechů se snažili první nebo poslední zvuk ignorovat. Navrhovali, aby bylo napříště použito paradigma 4IAX, v němž se referenční stimul ozve dvakrát, jednou před oběma neznámými zvuky. Bylo by tak možné změněný i původní stimul přímo a odděleně porovnat s referenčním. Na rozdíl od předchozího postřehu je toto spíš záležitostí pilotáže a kalibrace testu: položky v řazení 4IAX budou pravděpodobně obecně jednodušší než v pořadí AXB, a bude tedy potřeba upravit nápadnost manipulace, aby nedošlo ke stropovému efektu. Na druhou stranu ale taková změna řazení stimulů může vést i ke snížení frustrace respondentů, a to i navzdory tomu, že reálně jejich úspěšnost kvůli těžším stimulům neporoste – jen budou méně psychicky unaveni ze snahy střídavě ignorovat třetinu položky. V takovém případě by i tato změna byla přínosem.

S frustrací souvisí i to, že test byl přece jen poměrně dlouhý (30–60 min) a pro méně zdatné jedince těžký, takže někteří propadali lehké beznaději, byť byly úlohy celkem pestré. Potíž tkví v tom, že tito „neúspěšní“ respondenti jsou také potřeba, aby se projevila případná korelace. V budoucích experimentech by tedy možná bylo dobré zahrnout i některé výrazně jednodušší položky, aby se neztrácela motivace, či soustředit se na menší počet typů úloh (jen hudební a řečové; jen temporální manipulace; jen frekvenční apod.) a celkově test zkrátit.

Užitečnost požadavku ohodnotit u každé odpovědi stupeň subjektivní jistoty jsme si ukázali v odd. 5.2. Nicméně i zde by možná byly přínosné drobné úpravy, a to v

pojmenování tlačítek popisujících stupně jistoty. Jména použitá v našem experimentu byla „jsem si úplně jistý/á“, „jsem si celkem jistý/á“ a „spíš hádám“, přičemž první dvě úrovně jsme spojili do jedné, která reprezentuje všechny odpovědi, jež z našeho hlediska subjekt označil jako percepčně motivované (nenáhodné). Někteří respondenti ale měli po absolvování testu poznámky v duchu věty „Něco jsem tam slyšel/a, ale měl/a jsem pocit, že spíš hádám,“ které ukazují k tomu, že ačkoli jistou percepční motivaci měli, v některých případech zvolili nejhorší stupeň jistoty, což náš předpoklad narušuje.⁶² Problém tkví pravděpodobně ve slůvku „spíš“ a přílišné podobnosti prvních dvou možností; navrhneme zde tedy tentativně nové popisky pro tato tlačítka, která snad jejich intence vyjadřují explicitněji:

1. jsem si jistý/á
2. něco tam slyším, jen nevím přesně co
3. odpověď jsem vybral/a náhodně

Co se samotných řečových stimulů týče, lišila se u respondentů schopnost vědomé reflexe manipulací. Lingvisticky či přímo foneticky vzdělaní jedinci jejich podstatu často velmi přesně identifikovali, rozlišili temporální a intonační manipulace, dokonce komentovali i fakt že „fonologické“ (tj. významotvorné, překračující kategorie)⁶³ manipulace je mnohem snazší rozpoznat než manipulace „fonetické“ (rozdíly míry uvnitř kategorie). Jak jsme ovšem rozebrali v odd. 5.4, nebyla tato schopnost vědomé reflexe vyplývající z předmětu VŠ studia významným prediktorem úspěchu. Naopak respondent s druhým nejvyšším celkovým skóre popisoval charakter některých manipulací intonace v mlhavých termínech jako „větší důraz“, případně při vědomé reflexi dokonce domnělé místo manipulace lehce posunul, což mu ovšem nezabránilo rozdíly zaznamenat a v testu uspět. Citlivost na variace F0 je tedy pro řeč důležitá, ale při neškoleném vědomém posuzování máme tendenci uvažovat v jiných kategoriích než holá melodie, jako je důraz, příznakovost apod. Spíš než o čistě zvukové jde o jazykové kategorie, spjaté se sémantikou.

Pokud jde o hudební stimuly, bylo by zajímavé vyzkoušet, jakou by na úspěšnost v rozpoznání temporálních manipulací mělo vliv doplnění metronomu udávajícího takt do melodií. Bylo by také záhodno na základě percepčních experimentů vytvořit detailnější typologii manipulací s melodií podle jejich nápadnosti; v tomto výzkumu jsme při tvorbě stimulů jako heuristický údaj brali to (odd. 4.2), zda je kontura porušená či ne, ale celá problematika bude pravděpodobně složitější a provázaná i s rytmickým aspektem tónové sekvence. To už je ovšem práce spíše pro hudebního psychologa než fonetika.

⁶²Navzdory tomu, že během instruktáže experimentátor explicitně uváděl, že stupeň jistoty „spíš hádám“ odpovídá náhodné odpovědi.

⁶³Může jít např. o přesun intonačního centra, změnu kadence, ale i změnu trvání samohlásky, která již překročí hranici potřebnou pro změnu vnímané fonologické kategorie délky.

Smíšená část testu byla pro respondenty těžká v tom smyslu, že ne všichni, kteří prokázali percepční citlivost v řečové části, přišli „na kloub“ vztahu mezi větami a hudebními analogy (viz oslabená korelace v obr. 14). Navrhujeme zde tedy některá zjednodušení této úlohy, která by možná vedla k větší úspěšnosti při dekódování mezidoménového vztahu:

- ideálem zůstává lépe zmapovat vztahy mezi řečí a hudebními analogy, zaměřit sérii experimentů přímo na tuto problematiku (v rámci této práce na takový detailní průzkum nebyl prostor) a zjistit, jaká konkrétní převodní pravidla je potřeba respektovat, aby byl přenos do hudební domény co nejvěrnější.
- zajistit, aby analogy rozpětím F0 a barvou lépe odpovídaly svým řečovým protějškům (v této studii nebyl při volbě hudebních nástrojů či oktáv brán speciální ohled na to, že mluvčím vět je dospělý muž)
- jako distraktor použít hudební stylizaci odpovídající jinému řečovému stimulu, ne pouze v jednom bodě zmanipulovanou verzi správné stylizace. V takovém případě se oba nabízené analogy budou víc lišit, a úloha tak bude víc testovat cit pro celkovou podobnost. Pravděpodobně tedy nebude tak choulostivá na přesnost hudební stylizace a nebude vyžadovat, aby subjekt vůbec postřehl rozdíl mezi oběma stylizacemi, což je netriviální úkon, který v tomto typu úloh (na rozdíl od nesmíšených) vlastně testovat nechceme, ale přitom je vstupním požadavkem pro to, aby mohl subjekt fundovaně rozhodnout o tom, která stylizace je původní větě blíže.
- porovnávat větu ne s hudební stylizací, ale s variantami její resyntézy zbavené segmentálních charakteristik (zachová se jen intonační průběh). Zde ovšem nastává problém, že ačkoli jde také primárně o změny výšky v čase, status resyntézy jakožto hudby je pochybný.
- resyntéza by mohla být využita i naopak, namísto řečového stimulu, a hudební stylizace by byla modelovaná blíže podle ní, včetně mikrotonálních změn. Můžeme ovšem uplatnit stejnou námitku jako v předchozím bodě, jen v opačném gardu (do jaké míry je resyntéza ještě řečí?), o zvýšené technické náročnosti při tvorbě hudebních analogů (kvůli mikrotonalitě) nemluvě.

V každém případě by ale bylo možné v příštím výzkumu zahrnout resyntetizované stimuly jako samostatný typ či „mezidoménu“, která je takříkajíc na půli cesty mezi řečí a hudbou, a zahrnout ji do zkoumání korelací v diskriminační schopnosti. Lze uvažovat i o tom, že by se před samotnou resyntézou vytvořil z intonačního průběhu nejprve prozogram (viz Mertens 2004a a odd. 2.2.2), který se snaží modelovat naši percepci intonace, takže sekvenci zpřehlední a zdůrazní místa, která je potřeba v hudebním analogu zachytit.

Citelný dopad na výsledek našeho pokusu mohl mít fakt, že naše stimuly byly možná příliš dlouhé či syntakticky složité – moc mluvnických taktů, moc hudebních taktů a uvnitř nich komplikované vzájemné temporální i výškové vztahy mezi tóny. Velký vliv tak při

zpracování stimulů měly kapitální procesy („fonologická“ síta obou domén a mentální reprezentace použité při syntaktické integraci), které záleží na získaných dovednostech a zkušenosti v obou doménách, ne vrozených vlohách a obecné percepční citlivosti. To není samo o sobě špatná věc, ale testujeme vlastně něco víc než čistou percepční citlivost, a jak jsme již zmínili, jsou tak znevýhodněni zejm. ti respondenti, již mají menší zkušenost s hudební syntaxí (řeč používají aktivně všichni, hudbu jen někteří). Zde by se tedy potenciálně mohly skrývat rezervy experimentu, které vedly k rozostření korelací do širších pásem.

Kdybychom použili kratší, syntakticky jednodušší jednotky vyloženě zacílené na malé rozdíly, víc by vynikly schopnosti pro základní zpracování akustické informace prostřednictvím bazálních procesů. Tím by se potlačily rozdíly dané různým množstvím tréninku v syntaktické analýze hudby. Experiment tohoto typu, byť za účelem srovnání výkonu kongenitálních amúziků s referenčním vzorkem populace, popisuje Patel (2008, s. 230–231): všechny stimuly sestávaly ze sekvencí pěti tónů, vždy hraných na klavír, přičemž bylo dáno, že buď budou všechny stejné, nebo se čtvrtý tón v sekvenci (a nikdy žádný jiný) může lišit. V takto postaveném experimentu byli všichni posluchači z kontrolní skupiny (na rozdíl od amúziků) schopni velmi úspěšně zaznamenat i čtvrttónové variace, neboť test se skutečně zaměřil pouze na jejich percepční citlivost. V našem případě byly variace v absolutních číslech znatelně větší (3 nebo 4 ST, viz tab. 2), ale odváděla od nich pozornost mnohem složitější syntax i nejistota ohledně místa manipulace (respondent nevěděl, kde nastane), což vedlo k výrazně rozrůznějším výkonům. Je to jen další doklad toho, že právě postřehnutelný rozdíl je pojem relativní, silně závislý na konkrétní podobě úkolu, který posluchači plní, a jeho specifickém kontextu.

Schönová et al. (2004) zvolili opačnou cestu: provedli experiment, v němž se rozhodli vědomě cílit především na kapitální procesy.⁶⁴ Nešlo o porovnávání stimulů, nýbrž o identifikace porušení hudební či intonační kontury. Pokusné osoby tedy slyšely vždy jen jeden zvuk a měly rozhodnout, zda v jeho rámci dochází k nepřirozenému skoku v kontuře či nikoli. Tím pádem se autoři vyhnuli šumu, který v datech může způsobit fakt, že některé variace ve stimulech mohou být fonetického charakteru (změna stupně uvnitř kategorie) a jiné naopak fonologického (změna kategorie; platí jak pro řečové, tak pro hudební stimuly). V tomto případě byla totiž veškerá vybočení *a priori* fonologická, neb se vztahovala k očekáváním vycházejícím z předem osvojeného systému.

I tento přístup má ovšem svá úskalí. Zaprvé, skok v intonační kontuře nutně neznamená, že je kontura malformovaná, inkongruentní. Naopak může mít specifický pragmatický význam a je tedy otázkou, zda je vůbec oproti kontuře beze skoku fonologicky méně platná či dokonce neplatná. Zadruhé, jediný v článku uvedený příklad použité hudební melodie

⁶⁴Podobně Dankovičová et al. (2007) zkoumali korelaci hudebních schopností s jednou velmi specifickou získanou řečovou dovedností, a to intonační analýzou. Detaily viz odd. 2.2.4.

je nápěv z narozeninové písně „Hodně štěstí, zdraví“. Pokud pocházely všechny použité melodie z více či méně etablovaných písní, pak je nabíledni, že kromě hudební fonologie (citu pro tonalitu, případně pro typická tonální vybočení) vstupuje do rozhodování další nezanedbatelný faktor, a to obeznámenost s původní skladbou. Tomuto efektu jsme se použitím vlastních melodií a méně obvyklých církevních stupnic v naší práci vyhnuli.

Ještě další přístup by spočíval v tom, že by se řečové a hudební úlohy netvořily s takovým důrazem na paralelnost; místo toho by se použily specializované testy hudební a řečové dovednosti a porovnávali se výsledky v nich. V hudební doméně by bylo možné subjekty otestovat např. pomocí Montreal Battery of Evaluation of Amusia (MBEA, Peretz et al. 2003) či nějaké její upravené verze, nebo Wingova standardního psychometrického hudebního testu (1968). Další chytře navržené testy hudební dovednosti popisují Dankovičová et al. (2007, s. 183, 184 aj.); jde o hudební stimuly se spektrální obálkou modelovanou podle řeči, právě postřehnutelné rozdíly apod. Na straně řeči se nabízí např. výpočet verbálního IQ respondentů (viz např. Sadakata a Sekiyama 2011). Tím se ale dostáváme z fonetiky do oblasti psychologie. Nejde již o výzkum percepčního propojení, nýbrž možnosti sdílených obecných kognitivních dispozic, a je otázka, zda by vzhledem k vysokému počtu stupňů volnosti byl vůbec průkazný.

The image shows a musical score for three instruments: guitar (G), bass (B), and drums (D). The music is in 3/4 time and the key of D major (indicated by three sharps: F#, C#, G#). The guitar part consists of a series of eighth notes, many of which are beamed together in groups of two or four. Above the notes are circled numbers 2, indicating double stops or specific fingering. The bass part features a steady eighth-note pattern, with some notes beamed together. The drum part is represented by a single line with a series of '8' characters, likely indicating eighth notes or a specific drum pattern. The score is divided into two measures by a vertical bar line. The first measure contains the main melodic and rhythmic material, while the second measure shows a continuation of the bass line and a final chordal structure for the guitar.

Frank Zappa, „Joe's Garage“

Závěr

Cílem této práce bylo pomocí percepčního testu prověřit hypotézu, zda existuje korelace mezi percepční citlivostí na jemné frekvenční a temporální variace v řečových a hudebních stimulech. Kromě toho jsme zkoumali i možný vliv dalších faktorů na tuto citlivost, jmenovitě celkového vztahu respondentů k hudbě a jejich případného lingvisticko-fonetického vzdělání.

Teoretickou část jsme uvedli kapitolou o různých dimenzích, v nichž lze o vztahu hudby a řeči uvažovat, doplněnou o charakteristiku typických postupů výstavby zvukové struktury v obou doménách (kap. 1). V kap. 2 jsme následně podali přehled dosavadních existujících důkazů o spřízněnosti obou domén, ať už z oblasti evoluční biologie (odd. 2.1), neurověd (odd. 2.2.1) či psychofonetiky, tj. percepčních testů (odd. 2.2.4). Kromě toho jsme se také zabývali kvantifikací strukturních podobností mezi řečí a hudbou a jejich návazností na kulturní vlivy (odd. 2.2.2), jakož i problematikou právě postřehnutelných rozdílů ve frekvenční a temporální oblasti.

Experimentální část pak byla uvozena krátkým shrnutím použitého experimentálního paradigmatu, včetně přímočarých korelačních hypotéz, s nimiž jsme do výzkumu vstupovali (kap. 3). O konkrétních detailech našeho pojetí tohoto paradigmatu jsme obšírně poreferovali v kap. 4. Šlo zejm. o popis tvorby a pilotáže tří typů stimulů: řečových (věty o třech mluvních taktech), hudebních (melodie o třech dvoučtvrťových taktech) a smíšených (věty o třech mluvních taktech a jejich hudební analogy). V prvních dvou šlo o rozpoznání drobných temporálních a frekvenčních manipulací oproti referenčnímu stimulu, ve třetí pak o výběr hudebního analogu, který nejpřiléhavěji odpovídá původní větě.

Výsledky (kap. 5) ukazují, že korelace mezi výkony v jednotlivých částech testu skutečně existují, byť jsou do jisté míry rozostřené: většinou jde o středně silnou korelaci, někdy dokonce jen o korelaci nízkou (odd. 5.3). Pomocí lineárních modelů jsme se pak snažili integrovaně zkoumat vliv na úspěšnost v dané části testu jak výkonu v jeho ostatních částech, tak kategorických proměnných *hudebnost* a *VŠ*, které charakterizovaly respondentův celkový vztah k hudbě a to, zda navštěvoval VŠ v různé míře zaměřenou na studium jazyka. První z faktorů jistou roli hrál zejm. pro výkon v hudebním testu, druhý nikoli.

V diskusi (kap. 6) jsme se zaměřili na komplexnější analýzu výsledků, i s ohledem na poznatky shrnuté v teoretické části, především v odd. 2.2.4. Uvažovali jsme nad tím, co je vlastně vůbec percepční citlivost a zda si ji definujeme spíše jako bazální či kapitální proces. Naš experiment tyto dva aspekty propojoval a je možné, že tato interakce snížila ostrost výsledné korelace. Pokud v budoucím výzkumu přesněji zacílíme na bazální procesy (základní percepční citlivost), tj. úlohy, jež v hudební doméně neznevýhodňují ty respondenty, kteří se hudbě příliš nevěnují, ale ucho bystré mají, možná nalezneme korelaci i silnější. I s ohledem na tento výklad považujeme výsledky výzkumu jakožto spíše

ukazující na jisté provázání percepce temporálních a frekvenčních variací v řeči a hudbě, neboť menší ostrost zjištěného vztahu lze vysvětlit právě syntaktickou složitostí stimulů.

Mnoho otázek zůstává otevřených, a na některé z nich jsme se pokusit odpovědět ani nemohli. S Patelovou hypotézou o mezidoménovém sdílení mechanismů pro zpracování řeči a hudby, ne už tak mentálních reprezentací, je náš výzkum sice kompatibilní (viz závěr odd. 6.2), ale nijak ji nedokazuje ani nevyvrací, je v tomto ohledu pouze nepřímým důkazem. K debatě nad dalšími zajímavými aspekty problematiky pak nemůže přispět vůbec, ať už jde o hypotézu Abramse et al. (2011), že hudbu i řeč v mozku nezpracovávají výrazněji ohraničená centra, nýbrž propojená distribuovaná síť, či osvěžující pojetí Koelsche et al. (2002), kteří nabádají k většímu nadhledu a místo úzkého propojení mezi oběma doménami uvažují o obecnějších kognitivních mechanismech sdílených nejen hudbou a řečí, ale i dalšími mentálními úkony (obojí viz odd. 2.2.1).

Koneckonců i Leonard Bernstein nazval sérii svých harvardských přednášek, která v analogiích s lingvistickou teorií Noama Chomského zkoumala možnou existenci univerzálního hudebního jazyka, *The Unanswered Question* – nezodpovězená otázka. Na jejím konci se sice domníval, že jistou odpověď našel, ale pokračující diskreditace Chomského domněnek ve světle nových poznatků kognitivní lingvistiky a jejích spřízněných disciplín ukazuje, že bude nutné hledat dál.

Bibliografie

- ABERCROMBIE, David, 1965. *Studies in phonetics and linguistics*. London: OUP.
- ABLA, Dilshat, Kentaro KATAHIRA a Kazuo OKANOYA, 2008. On-line assessment of statistical learning by event-related potentials. *J Cognitive Neurosci*. roč. 20, č. 6, s. 952–964.
- ABRAMS, Daniel A., Anjali BHATARA, Srikanth RYALI, Evan BALABAN, Daniel J. LEVITIN a Vinod MENON, 2011. Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cerebral Cortex*. roč. 21, č., s. 1507–1518.
- BERWICK, Robert C., Kazuo OKANOYA, Gabriel J.L. BECKERS a Johan J. BOLHUIS, 2011. Songs to syntax: the linguistics of birdsong. *Trends in Cognitive Sciences*. March., roč. 15, č. 3, s. 113–121.
- BOERSMA, Paul a David WEENINK, 2014. *Praat: doing phonetics by computer [počítačový program]*. v5.3.73 [online]. 2014. Dostupné z: <http://www.praat.org/>
- BOLHUIS, Johan J., Kazuo OKANOYA a Constance SCHARFF, 2010. Twitter evolution: converging mechanisms in birdsong and human speech. *Nat Rev Neurosci*. November., roč. 11, č., s. 747–759.
- BRANCAZIO, L., C.T. BEST a C.A. FOWLER, 2006. Visual influences on perception of speech and nonspeech vocal-tract events. *Language and Speech*. roč. 49, č. 1, s. 21–53.
- BROWN, Steven, 2000. The „musilanguage” model of music evolution. In: Nils Lennart WALLIN, Björn MERKER a Steven BROWN, ed. *The origins of music*. Cambridge, MA: MIT Press, s. 271–300.
- DANKOVIČOVÁ, Jana, Jill HOUSE, Anna CROOKS a Katie JONES, 2007. The relationship between musical skills, music training, and intonation analysis skills. *Language and Speech*. B.m.: Kingston Press Ltd., roč. 50, č. 2, s. 177–225.
- DARWIN, Charles, 2006. *O původu člověka*. Přel. Josef WOLF a Zora WOLFOVÁ. Praha: Academia.
- DELLWO, Volker, 2006. Rhythm and speech rate: A variation coefficient for C. In: Pawel KARNOWSKI a Imre SZIGETI, ed. *Language and language-processing*. Frankfurt am Main: Peter Lang, s. 231–241.
- DELLWO, Volker a Petra WAGNER, 2003. Relations between language rhythm and speech rate. In: *ICPhS 2003: Proceedings of the xVth international congress of phonetic sciences*. Barcelona: UAB, s. 471–474.
- DEUTSCH, D., T. HENTHORN, E. MARVIN a H.-S. XU, 2006. Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *J Acoust Soc Am*. roč. 119, č., s. 719–722.
- FEDORENKO, Evelina, Aniruddh D. PATEL, Daniel CASSANTO, Jonathan WINA-

WER a Edward GIBSON, 2009. Structural integration in language and music: Evidence for a shared system. *Memory & Cognition*. roč. 37, č. 1, s. 1–9.

FODOR, J.A. a T.G. BEVER, 1965. The psychological reality of linguistic segments. *J Verb Learn Verb Behav*. October., roč. 4, č. 5, s. 414–420.

FORDE THOMPSON, William, Manuela M. MARIN a Lauren STEWART, 2012. Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *PNAS*. roč. 109, č. 46, s. 19027–19032.

GARRATT, James, 2010. *Music, culture and social reform in the age of wagner*. Cambridge: CUP.

GRABE, Esther a Ee Ling LOW, 2002. Durational variability in speech and the rhythm class hypothesis. In: Carlos GUSSENHOVEN a Natasha WARNER, ed. *Laboratory phonology 7*. Berlin, New York: Mouton de Gruyter, s. 515–546.

HANNON, Erin E., 2009. Perceiving speech rhythm in music: Listeners classify instrumental songs according to language of origin. *Cognition*. č. 111, s. 403–409.

HAUSER, Marc D., Noam CHOMSKY a William Tecumseh FITCH, 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science*. November., roč. 298, č., s. 1569–1579.

HELMHOLTZ, Hermann L. F., 1885. *On the sensations of tone as a physiological basis for the theory of music*. Přel. Alexander J. ELLIS. London, New York: Longmans, Green,; Co.

JACKENDOFF, Ray, 2009. Parallels and nonparallels between language and music. *Music Perception*. roč. 26, č. 3, s. 195–204.

JIANG, Cunmei, Jeff P. HAMM, Vanessa K. LIM, Ian J. KIRK a Yufang YANG, 2010. Processing melodic countour and speech intonation in congenital amusics with mandarin chinese. *Neuropsychologia*. B.m.: Elsevier Ltd., roč. 48, č., s. 2630–2639.

KAGAWA, Hiroko, Hiroko YAMADA, Ruey-shing LIN, Taku MIZUTA, Toshikazu HASEGAWA a Kazuo OKANOYA, 2012. Ecological correlates of song complexity in white-rumped munias: The implication of relaxation of selection as a cause for signal variation in birdsong. *Interaction Studies*. B.m.: John Benjamins Publishing Company, roč. 13, č. 2, s. 263–284.

KOBAYASHI, K., H. UNO a K. OKANOYA, 2001. Partial lesions in the anterior forebrain pathway affect song production in adult bengalese finches. *Neuroreport*. roč. 12, č., s. 353–358.

KOELSCH, Stefan, 2011. Towards a neural basis of processing musical semantics. *Phys Life Rev*. roč. 8, č., s. 89–105.

KOELSCH, Stefan, Thomas C. GUNTER, Yves V. CRAMON, Stefan ZYSSET, Gabriele LOHMANN a Angela D. FRIEDERICI, 2002. Bach speaks: A cortical „language-Network” serves the processing of music. *NeuroImage*. B.m.: Elsevier Science (USA), roč.

17, č., s. 956–966.

KOELSCH, Stefan, Elisabeth KASPER, Daniela SAMMLER, Katrin SCHULZE, Thomas GUNTER a Angela D. FRIEDERICI, 2004. Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience*. roč. 7, č. 3, s. 302–307.

KROOS, Christian a Katherine HOGAN, 2009. Visual influence on auditory perception: Is speech special? In: B.-J. THEOBALD a R.W. HARVEY, ed. *AVSP 2009 – international conference on audio-visual speech processing university of east anglia, norwich, uK; september 10–13, 2009*. B.m.: ISCA, s. 70–75.

LABOV, William, 1963. The social motivation of a sound change. *Word*. roč. 19, č., s. 273–309.

LEE, Chao-Yang a Yuh-Fang LEE, 2010. Perception of musical pitch and lexical tones by mandarin-speaking musicians. *J Acoust Soc Am*. January., roč. 127, č. 1, s. 481–490.

LEVITIN, Daniel J. a Vinod MENON, 2003. Musical structure is processed in „language” areas of the brain: a possible role for brodmann area 47 in temporal coherence. *NeuroImage*. B.m.: Elsevier Inc., roč. 20, č., s. 2142–2152.

LEWIS, Jerome, 2012. Response to richard widdess: Music, meaning and culture. *Empirical Musicology Review*. roč. 7, č. 1–2, s. 98–101.

MERTENS, Piet, 2004a. The prosogram: Semi-automatic transcription of prosody based on a tonal perception model. In: *Proceedings of speech prosody 2004*. s. 23–26.

MERTENS, Piet, 2004b. Un outil pour la transcription de la prosodie dans les corpus oraux. *Traitement Automatique des langues*. roč. 45, č. 2, s. 109–130.

MOK, P.K. Peggy a Donghui ZUO, 2012. The separation between music and speech: Evidence from the perception of cantonese tones. *J Acoust Soc Am*. October., roč. 132, č. 4, s. 2711–2720.

NIETZSCHE, Friedrich, 2008. *Zrození tragédie*. Přel. Otokar FISCHER. Praha: Vyšehrad.

NOOTEBOOM, Sieb, 1999. The Prosody of Speech: Melody and Rhythm. In: William J. HARDCASTLE a John LAVER, ed. *The handbook of phonetic sciences*. Blackwell Reference Online: Blackwell Publishing, Blackwell handbooks in linguistics.

OKANOYA, Kazuo, 2004. The bengalese finch: A window on the behavioral neurobiology of birdsong syntax. *Ann NY Acad Sci*. roč. 1016, č., s. 724–735.

OKANOYA, Kazuo, 2007. Language evolution and an emergent property. *Curr Opin Neurobiol*. roč. 17, č., s. 271–276.

OKANOYA, Kazuo, 2013. *Bird song as a musical protolanguage*. 2013. B.m.: Příspěvek přednesený 17. 9. 2013 na *International Summer School on Agent-based Models of Creativity* v Cortoně, Itálie; video z přednášky je dostupné na <<http://www.youtube.com/watch?v=NlfVeAE1Eo8>>.

OKANOYA, Kazuo a Bjorn MERKER, 2006. Neural substrates for string-context

mutual segmentation: a path to human language. In: Caroline LYON, Christopher L. NEHANIV a Angelo CANGELOSI, ed. *Emergence of communication and language*. London: Springer-Verlag, s. 421–434.

OKANOYA, Kazuo, Sayaka HIHARA, Naoko TOKIMOTO, Yasuko TOBARI a Atsushi IRIKI, 2007. Complex vocal behavior and cortical-medullar projection. *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer-Verlag, roč. 3609, č., s. 362–367.

OUDEYER, Pierre-Yves, 2006. *Self-organization in the evolution of speech*. Přel. James R. HURFORD. Oxford, New York: OUP.

PATEL, Aniruddh D., 2003. Language, music, syntax, and the brain. *Nature Neuroscience*. roč. 6, č., s. 674–681.

PATEL, Aniruddh D., 2005. The relationship of music to the melody of speech and to syntactic processing disorders in aphasia. *Ann NY Acad Sci*. roč. 1060, č., s. 59–70.

PATEL, Aniruddh D., 2008. *Music, language, and the brain*. New York: OUP.

PATEL, Aniruddh D. a Joseph R. DANIELE, 2003. An empirical comparison of rhythm in language and music. *Cognition*. č. 87, s. B35–B45.

PATEL, Aniruddh D., John R. IVERSEN a Jason C. ROSENBERG, 2006. Comparing the rhythm and melody of speech and music: The case of british english and french. *J Acoust Soc Am*. roč. 119, č. 5, s. 3034–3047.

PERETZ, Isabelle, Anne Sophie CHAMPOD a Krista HYDE, 2003. Varieties of musical disorders: The Montreal Battery of Evaluation of Amusia. *Ann NY Acad Sci*. roč. 999, č., s. 58–75.

PIKE, Kenneth, 1945. *The intonation of american english*. Ann Arbor, MI: University of Michigan Press.

PINKER, Steven, 2005. So how *does* the mind work? *Mind & Language*. February., roč. 20, č. 1, s. 1–24.

POLS, Louis C.W., 1999. Flexible, robust and efficient human speech processing versus present-day speech technology. In: *ICPhS 1999: Proceedings of the xIVth international congress of phonetic sciences*. San Francisco: University of California, s. 9–17.

R CORE TEAM, 2014. *R: A language and environment for statistical computing* [online]. Vienna, Austria: R Foundation for Statistical Computing. Dostupné z: <http://www.R-project.org>

RAMUS, Franck, Marina NESPOR a Jacques MEHLER, 1999. Correlates of linguistic rhythm in the speech signal. *Cognition*. B.m.: Elsevier Science B.V., roč. 73, č., s. 265–292.

ROACH, Peter, 1982. On the distinction between „stress-timed” and „syllable-timed” languages. In: David CRYSTAL, ed. *Linguistic controversies*. London: MIT Press, s. 73–79.

SADAKATA, Makiko a Kaoru SEKIYAMA, 2011. Enhanced perception of various linguistic features by musicians: A cross-linguistic study. *Acta Psychologica*. B.m.: Elsevier

B.V., roč. 138, č., s. 1–10.

SCHÖN, Daniele, Cyrille MAGNE a Mireille BESSON, 2004. The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*. roč. 41, č., s. 341–349.

SLEVC, L. Robert a Akira MIYAKE, 2006. Individual differences in second language proficiency: Does musical ability matter? *Psychological Science*. roč. 17, č. 8, s. 675–681.

SLEVC, L. Robert a Aniruddh D. PATEL, 2011. Meaning in music and language: Three key differences: Comment on „towards a neural basis of processing musical semantics” by stefan koelsch. *Physics of Life Reviews*. roč. 8, č., s. 110–111.

STEELS, Luc, 2011. Modeling the cultural evolution of language. *Physics of Life Reviews*. roč. 8, č., s. 339–356.

SUGE, Rie a Kazuo OKANOYA, 2010. Perceptual chunking in the self-perceived songs of bengalese finches (*lonchura striata* var. *domestica*). *Anim Cog*. B.m.: Springer-Verlag, roč. 13, č., s. 515–523.

'T HART, J.T., R. COLLIER a A. COHEN, 1990. *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge: CUP.

TAKAHASHI, Miki, Hiroko YAMADA a Kazuo OKANOYA, 2010. Statistical and prosodic cues for song segmentation learning by bengalese finches (*lonchura striata* var. *domestica*). *Ethology*. B.m.: Blackwell Verlag GmbH, roč. 116, č., s. 481–489.

UEKITA, Tomoko a Kazuo OKANOYA, 2011. Hippocampus lesions induced deficits in social and spatial recognition in *octodon degus*. *Behav Brain Res*. roč. 219, č., s. 302–309.

ÚČNK FF UK, 2010. *Český národní korpus – SYN2010* [online]. 2010. Dostupné z: <http://www.korpus.cz>

VOLÍN, Jan, 2007. *Statistické metody ve fonetickém výzkumu*. Praha: Nakladatelství EPOCH.

VUUST, Peter, Andreas ROEPSTORFF, Mikkel WALLENTIN, Kim MOURIDSEN a Leif ØSTERGAARD, 2006. It don't mean a thing... keeping the rhythm during poly-rhythmic tension, activates language areas (bA47). *NeuroImage*. roč. 31, č., s. 832–841.

VUUST, Peter, Mikkel WALLENTIN, Kim MOURIDSEN, Leif ØSTERGAARD a Andreas ROEPSTORFF, 2011. Tapping polyrhythms in music activates language areas. *Neuroscience Letters*. roč. 494, č., s. 211–216.

WING, H.D., 1968. *Tests of musical ability and appreciation: An investigation into the measurement, distribution, and development of musical capacity*. second. London: CUP.