

Filozofická fakulta Univerzity Karlovy

Fonetický ústav

BAKALÁŘSKÁ PRÁCE

Zuzana Prokopová

VLIV SEGMENTÁLNÍHO OKOLÍ NA VNÍMÁNÍ NAZÁLNÍCH KONSONANTŮ

The effect of segmental context on the perception of nasal
consonants

Praha, 2017

Vedoucí práce:

Ing. Tomáš Bořil, Ph.D.

Poděkování

Upřímně děkuji vedoucímu práce Tomáši Bořilovi za trpělivost a za množství rad a podnětů, jejichž možné využití dalece přesahuje rámec této práce. Dále jsem hluboce vděčná Radku Ocelákovi, který svou pílí vytvářel prostředí motivující k soustředěné práci.

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze dne 9. srpna 2017.

Podpis:

Abstrakt

Práce zkoumá vliv segmentálního okolí na percepci místa artikulace nazálních konsonantů. Z nahrávek jednoho mluvčího byly vyříznuty dvojslabičné sekvence končící slabikami [ne] či [na:] a hláska předcházející nazále byla obměňována jednak nahrazováním celé první slabiky slabikou jinou, avšak pocházející z fonémicky totožného kontextu, jednak manipulacemi formantových trajektorií. 15 původních, 15 lepených a 5 manipulovaných pseudoslov bylo předloženo 32 posluchačům, jejichž úkolem bylo zvolit mezi variantou pseudoslova obsahující [n] a variantou obsahující [m] tu, jež odpovídala slyšenému stimulu. U šesti lepených a jedné manipulované položky měla obměna první slabiky za následek statisticky významný posun průměrného hodnocení: ve zdrojových stimulech bylo jasně identifikováno [n], kdežto u těchto položek bylo hodnocení přinejmenším nejednoznačné. Tento náález hovoří proti existenci invariantního akustického korelátu místa artikulace v nazále či některé z okolních hlásek a podtrhuje kontextuální závislost vnímání segmentů. Výsledek práce podpořený navazujícím výzkumem by mohl přispět ke zkvalitnění výstupů konkatenční syntézy řeči.

Klíčová slova: nazální konsonanty, místo artikulace, vliv segmentálního okolí, percepcie řeči

Abstract

In this thesis, the effect of segmental context on the perception of nasal consonants (namely their place of articulation) is investigated. Disyllabic sequences ending with syllables [ne] or [na:] were extracted from one speaker's utterances and the vowel preceding the nasal consonant was varied either by replacing the first syllable with a different one, which however came from the same phonemic context, or by manipulation of formant trajectories. 15 original, 15 spliced and 5 manipulated pseudo-words were presented to 32 listeners, whose task was to choose from two options, one containing [n] and the other [m], the one that matched the heard stimulus. In case of six spliced and one manipulated stimuli, there was a significant shift of the average evaluation: while the original stimuli were clearly identified as containing [n], evaluation of the altered stimuli was at least equivocal. This finding is inconsistent with the idea of an invariant acoustic cue for place of articulation in the nasal consonant itself or in one of the adjoining segments, and emphasises the contextual dependence of our perception of segments. The result may be of importance for concatenative speech synthesis.

Keywords: nasal consonants, place of articulation, effect of segmental context, perception of speech

Obsah

1 Úvod	7
1.1 Motivace práce	7
1.2 Cíl a postup práce	8
2 Shrnutí předchozího výzkumu	9
2.1 Percepce řeči	9
2.1.1 Teorie percepce řeči	10
2.1.2 Segmentální přístup a jeho limity	11
2.1.3 Vliv kontextu na percepci segmentů	13
2.2 Percepce místa artikulace okluziv	16
2.3 Nazály	19
2.3.1 Akustické vlastnosti	19
2.3.2 Percepce místa artikulace	21
3 Metoda	28
3.1 Zvukový materiál	28
3.2 Tvorba stimulů	28
3.3 Percepční test	31
3.4 Respondenti	32
3.5 Zpracování dat	33
4 Výsledky	34
5 Diskuse	37
5.1 Interpretace výsledků	37
5.2 Návrh navazujícího výzkumu	40
5.3 Možný přínos uvedených výsledků	41
6 Závěr	42
7 Přílohy	46

1 Úvod

1.1 Motivace práce

Vznik této práce byl podnícen překvapivým úkazem, který se objevil při testování řečového syntetizéru vytvořeného v rámci Katedry kybernetiky Fakulty aplikovaných věd Západočeské univerzity v Plzni (KKY)¹. Syntetizované slovo „mlýneček“ znělo několika posluchačům i při opakovaném a pozorném poslechu jako „mlýmeček“. Že se jedná o úkaz hodný pozornosti, a nikoliv o selhání syntézy, vyvstane z následujícího odstavce, který ve stručnosti představí princip fungování plzeňského syntetizéru.

Jedná se o syntézu konkatenáční (tedy „zřetězovací“), která nové promluvy vytváří řetězením malých segmentů pocházejících z rozsáhlého korpusu přirozených promluv nahraného jedním mluvčím (který tímto propůjčuje syntéze svůj hlas). Nahrávky jsou segmentovány na úseky odpovídající hláskám, zvané fony (respektive trifony – ke každému fonu totiž zůstává připojena informace o hláskách v bezprostředním okolí), z nichž jsou následně vytvořeny difony prostým posunutím hranice každého segmentu do jeho poloviny. Vzniknou tedy jednotky typu [s_a], které obsahují vždy polovinu sousedících hlásek (zde druhou polovinu [s] a první polovinu [a]) a k tomu informaci, která hláska původně předcházela [s] a která následovala po [a]. Z těchto jednotek pak algoritmus podle značného množství parametrů vybírá tu nejvhodnější pro nový kontext. Chyby, kterých se syntetizér dopouští na segmentální úrovni, pramení nejčastěji z nesprávného označení hlásky v korpusu (segmentace a labelování probíhá vzhledem k velkému rozsahu korpusu automaticky) nebo z nedostatků ve fonetické transkripci ortografické podoby promluvy. Ani jedno však nebyl případ „mlýmečku“ – foném /n/ byl transkribován správně a použité difony pocházely ze správně nasegmentovaných slov skutečně obsahujících jeho realizaci (konkrétně ze slov „Jelínka“ a „Janečka“), přičemž v těchto zdrojových slovech byl daný segment zcela bez potíží vnímán jako [n] (jak si čtenář může sám ověřit v Příloze 1). Tento výchozí případ je ovšem specifický tím, že výsledný segment [n] byl slepen ze dvou různých alofonů fonému /n/, totiž [n] a [ŋ]. Tyto alofony jsou sice

¹Jedná se o systém ARTIC (*Artificial Talker in Czech*), viz Matoušek, Tihelka a Romportl (2006).

v korpusu odlišeny, ale pro algoritmus syntézy zjevně nemá jejich rozdílnost velkou (či dokonce žádnou) váhu, neboť slovo „Jelínka“ bylo zvoleno jako nejvhodnější kandidát pro tento kontext. V tomto případě by se tedy dal výsledný percepční dojem vysvětlit touto nesrovnalostí; jev se však opakoval i po nahrazení zdrojového slova „Jelínka“ slovem obsahujícím alofon [n] (viz opět Příloha 1).

1.2 Cíl a postup práce

Cílem této práce je otestovat hypotézu, že výše popsaná změna v percepci nazálního konsonantu je způsobena jeho segmentálním okolím. Stěžejní je tudíž vyloučit možnost, že popsaný jev je způsobován čistě diskontinuitou uvnitř samotné nazály, jež je nevyhnutelným důsledkem konkatenací syntézy operující s difony. Práce se tedy pokusí ověřit na skupině nezavěšených posluchačů, že tento jev nastává i v případě, že nazální konsonant, jehož vnímání sledujeme, zůstává neporušen – tedy že pouhá změna segmentálního okolí může způsobit, že segment vnímaný ve svém původním kontextu jako [n] je vnímán jako [m]², nebo je alespoň jeho identifikace jakožto [n] výrazně narušena.

Struktura práce je následující: Druhá kapitola je věnována teoretickému pozadí zkoumaného jevu. Po krátkém uvedení do problematiky percepce řeči obecně (2.1) jsou shrnuty klíčové momenty dosavadního výzkumu percepce místa artikulace okluzivních konsonantů (2.2). Zvýšená pozornost je ovšem věnována nazálám: jsou představeny jejich akustické vlastnosti (2.3.1) a jejich dosud zdokumentovaný vztah k percepci místa artikulace (2.3.2). Třetí kapitola seznámí čtenáře s metodou experimentu, jenž tvoří jádro práce. Je zde důkladně popsán použitý zvukový materiál (3.1) a postup při tvorbě stimulů využitých v percepčním testu (3.2). Oddíly 3.3 a 3.4 pak obsahují informace o vytváření samotného testu, jeho průběhu a o zúčastněných posluchačích. Zpracování výsledků je popsáno v oddílu 3.5 a výsledkům samým je věnována celá kapitola čtvrtá (4). Náplní páté kapitoly je diskuse, jež zahrnuje zaprvé interpretaci výsledků (5.1), zadruhé návrh směřování navazujícího výzkumu (5.2) a zatřetí nastínění možného přínosu této práce, zejména s ohledem na konkatenací syntézu řeči (5.3).

²Třetí český nazální foném /ɲ/ tato práce nezkoumá; zaprvé by takové rozšíření výzkumné otázky příliš zvětšilo její rozsah, zadruhé se i většina dosavadních studií týkala pouze uvedeného dvojice, takže toto omezení usnadní propojení práce s předchozím výzkumem.

2 Shrnutí předchozího výzkumu

2.1 Percepce řeči

Percepce řeči je značně širokým výzkumným polem, na němž se setkává hned několik vědních disciplín (kromě fonetiky například akustika, neurologie či kognitivní psychologie). Procesy, které transformují výchylky tlaku vzduchu působené mluvčím v něco, čemu posluchač *rozumí* (to jest identifikuje signál jako řeč a úspěšně jej uvede do vztahu se svou dosavadní lingvistickou a sociální zkušeností), probíhají tak rychle a samozřejmě, že jim v běžném životě téměř nevěnujeme pozornost; tím obtížnější je pak odhalit jejich přesné fungování. Popis percepce je ostatně složitý i tehdy, když nás zajímají akusticky výrazně jednodušší stimuly, než je řeč. Vztah mezi klíčovými fyzikálními dimenzemi zvuku (frekvencí, amplitudou, trváním a spektrálním složením) a jejich psychoakustickými koreláty (výškou, hlasitostí, délkou a barvou/kvalitou) není vůbec přímočarý a je stále předmětem výzkumu (Skarnitzl, Šturm a Volín, 2016, kap. 3). Lidská řeč je nadto signál velmi komplexní a mnohvrstevnatý – uvažme, kolik informací je najednou přítomno třeba jen ve zvuku slov „Tak běž!“. Rozpoznáváme tento zvuk jako řeč, přesněji jako česká slova; odlišujeme je od spousty dalších slov jako třeba „pak běž“ či „tak lež“; poznáme, že je řekla žena; co víc – poznáme, že je vykřikla, a také že ten výkřik nebyl rozzlobený, nýbrž vzrušený jako třeba při dostizích; mohli bychom dokonce i rozeznat, že ta žena je mladá cizinka... A to vše je náš percepční aparát schopen získat z necelé sekundy akustického signálu, navíc třeba smíšeného s okolním ruchem. Není proto divu, že percepce řeči je fenoménem stejnou měrou složitým jako fascinujícím.

Tato práce se soustředí na identifikaci segmentů za specifických podmínek – setrvává tedy v oblasti patrně nejsnáze vědecky uchopitelné, a proto také nejprobadanější. I v jejím rámci lze ovšem nalézt různé teoretické přístupy hledající odpovědi na stále nevyjasněné základní otázky; zejména na tzv. problém invariance. Ten odkazuje ke skutečnosti, že lidé jsou schopni v řeči identifikovat poměrně malé množství opakujících se lingvistických jednotek, ačkoliv variabilita řečového signálu je nesmírná – vyřešit problém invariance by znamenalo nalézt neměnné koreláty těchto jednotek, které při percepci využíváme (anebo uchopit percepci zcela jinak). Výzkum na tomto poli je rozvětvený a bohatý (Pisoni a Remez, 2005); zde uvedeme

tři hlavní přístupy k problému invariance a k percepci řeči obecně, tak jak je popisují práce Diehl, Lotto a Holt (2004) a Mitterer a Cutler (2006).

2.1.1 Teorie percepce řeči

Motorická teorie (*Motor Theory*), podníčená výzkumy provedenými v 50. a 60. letech v *Haskins Laboratories* v New Yorku, o nichž bude řeč v sekci 2.2, předpokládá velmi silný vztah percepce řeči k její produkci. Zabývá se percepcí řečových jednotek na úrovni hlásek a tyto jednotky podle ní posluchač nerozpoznává přímo z akustických vlastností signálu, nýbrž na základě těchto vlastností rekonstruuje pohyby řečových orgánů zamýšlené mluvčím. Vyhodnocovány jsou tedy v posledku namísto samotného přijímaného signálu předpokládané neuromotorické příkazy, které jej způsobily. K takto abstraktní rovině se motorická teorie uchyluje proto, že až tam máme podle jejích autorů co do činění s kontextuálně nezávislými korelátami lingvistických jednotek. Že akustický signál samotný takovéto korelátu neobsahuje, se zdá dosvědčovat už pouhý pohled na spektrogramy různých realizací téhož fonému – akusticky si tyto segmenty mohou být až překvapivě málo podobné. Jednou z příčin této variability je fyziologie našeho produkčního aparátu (při jeho mechanických omezeních a při rychlosti, jakou řeč produkujeme, od něj ani nelze tvorbu invariantních korelátů fonémů očekávat – viz níže výklad o koartikulaci), a proto se motorická teorie nezastavuje u skutečně prováděných artikulačních pohybů a předpokládá rekonstrukci pohybů intendovaných. Ruku v ruce s tímto předpokladem jde nevyhnutelně kladení jakéhosi „pokusného syntetizéru“ do našeho percepčního aparátu – ten slouží k nalezení neuromotorického příkazu, který s největší pravděpodobností stojí za daným segmentem akustického signálu. Podle motorické teorie je tento nástroj vlastní pouze člověku jakožto bytosti disponující řečí.

Teorie přímé percepce (*Direct-Perception Theory*) vychází z předpokladu, že v akustickém signálu je obsažen dostatek jednoznačně interpretovatelných informací o tom, jak daný zvuk vznikl (nemusí se zde jednat pouze o zvuky řečové – teorie přímé percepce se týká vnímání obecně). Ty nám zajišťují přímý přístup k učiněným artikulačním pohybům. Tato teorie tedy sdílí s motorickou teorií základní předpoklad úzkého sepětí percepce s produkcí, avšak odmítá postulovat unikátní, pouze člověku náležející nástroj rekonstruuující neuromotorické příkazy. Musí se však tedy

vypořádat s problémem invariance jinak, a také to činí – její odpovědí na tento problém je hypotéza „vektorové analýzy“ řečového signálu (Fowler a Smith, 1986). Ta se nedívá na řeč jako na řadu po sobě jdoucích segmentů, které se vzájemně ovlivňují, takže každý z nich v sobě nese informace i o segmentech okolních (takto je s řečí nakládáno při běžné fonetické segmentaci, aplikované i v této práci), nýbrž jako na segmenty, jež se v čase překrývají a každý z nich obsahuje pouze informace o sobě samém. Vše, co mají na svědomí segmenty okolní, je odfiltrováno, přičemž díky možnosti časového překryvu se žádná informace z původního signálu neztratí. Podle této hypotézy se tedy jednotlivé segmenty řeči (a artikulační úkony, které je produkují) v přísném slova smyslu navzájem neovlivňují, nýbrž se pouze překrývají – a percepční aparát je schopen „rozebrat“ takto strukturovaný signál na jednotlivé komponenty, které pak odpovídají lingvistickým jednotkám. Takovéto pojetí struktury řečového signálu umožňuje teorii přímé percepce obejít se bez překroku k abstraktnějším entitám typu neuromotorických příkazů.

Přístupy založené na obecných procesech slyšení a učení (*General Auditory and Learning Approaches*) tvoří širší teoretický rámec sdružující dílčí teorie, které na rozdíl od obou předchozích popisují percepci řeči nezávisle na její produkci a také odmítají postulování mechanismů a procesů specifických pouze pro percepci řeči. Tento pohled nahlíží na řeč pouze jako na jeden zvukový signál mezi mnoha jinými, se kterým se lidé vyrovnávají pomocí týchž mechanismů jako s jakýmkoliv dalším. Důležitá role zde náleží mechanismu percepčního učení – obecné auditivní procesy umožní získávat ze signálu informace, které se posluchač postupně naučí využívat jako vodítka pro rozpoznání opakujících se (například lingvisticky relevantních) zvuků. Ve prospěch tohoto přístupu hovoří jednak studie dokládající, že percepcce neřečových stimulů vykazuje stejné vlastnosti jako percepcce řeči, jednak pozorování obdobných percepčních procesů u jiných živočišných druhů.

2.1.2 Segmentální přístup a jeho limity

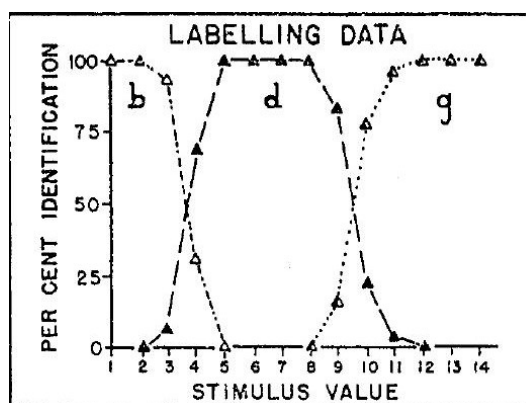
Společným znakem výše popsaných teorií (přinejmenším prvních dvou – naposled popsaný přístup je natolik obecný, že neumožňuje příliš mnoho tvrzení týkajících se ho jako celku) je, že řeč a její percepci pojmají segmentálně, jakkoliv výrazně se v pojetí segmentů od sebe liší. Jelikož základním rámcem i této práce je segmentální

přístup, je třeba zmínit jeho limity.

Dělení akustického kontinua, jímž mluvená řeč je, na segmenty odpovídající hláskám je sice běžná praxe a má z hlediska výzkumu jisté opodstatnění (Mitterer a Cutler, 2006, s. 770), neodráží však příliš věrně ani produkci, ani percepci řeči (Skarnitzl et al., 2016, s. 108–109), jak ostatně dokládá i palčivost problému invariance (viz 2.1). Při produkci zcela nevyhnutelně dochází ke koartikulaci (Skarnitzl et al., 2016, s. 71–73; Farnetani a Recasens, 2010, s. 316–320) – mísení či překryvům artikulačních a fonačních úkonů, které by v zájmu bezproblémové segmentace musely být přísně odděleny. Nazalizace vokálu sousedícího s nazálním konsonantem (vznikající kvůli omezené rychlosti pohybu měkkého patra, které by mělo být v době vyslovování vokálu zdvižené a v době vyslovování nazály spuštěné) může být jedním příkladem; vliv konsonantu na formantové trajektorie sousedních vokálů, o němž bude podrobněji pojednáno v sekci 2.2, dalším. Zmíněné teorie percepcie řeči se sice různým způsobem snaží se s touto skutečností vyrovnávat (a do nějaké míry jistě zdařile), ale i tak je třeba vzít na vědomí, že řada novějších výzkumů ukazuje samotné dělení řeči na hláskám odpovídající segmenty jako umělé (viz opět Skarnitzl et al., 2016, s. 109).

Ačkoliv tedy experimenty pracující se segmenty snad mohou přinášet užitečné poznatky, je při jejich interpretaci třeba mít na paměti povahu přirozené řeči; alespoň tyto dva, zde zkratkovitě vyjádřené, rysy: Zaprvé že akustické vlastnosti segmentů jsou variabilní například v závislosti na svém kontextu (viz výše zmíněná koartikulace) či na mluvčím (viz např. Johnson, 2005); a zadruhé že tytéž faktory (a ovšem nejen ony) ovlivňují i to, jak segmenty vnímáme, byť by byly akusticky totožné. Jelikož druhý uvedený bod (konkrétně vliv kontextu na percepci segmentů) je pro tuto práci stěžejní, bude rozveden v následujícím oddíle.

Avšak ještě než se zaměříme na vliv okolí, je třeba představit jednu podstatnou a dnes již nespornou vlastnost našeho vnímání segmentů – totiž jeho kategoriálnost (viz klasickou studii Liberman, Harris, Hoffman a Griffith, 1957 či pro shrnutí Mitterer a Cutler, 2006). Měníme-li akustický parametr jistého zvuku po malých krocích tak, že tím přecházíme od typické akustické podoby jednoho fonému (například /b/) k typické podobě fonému jiného (například /d/), a požádáme-li pak posluchače o identifikaci daných zvuků, zjistíme, že percepcie nekopíruje plynulost akustického



Obrázek 1: Ilustrace kategoriální percepce (převzato ze studie Liberman et al., 1957). Na ose x jsou vyneseny indexy jednotlivých položek, lišících se směřováním tranzientu F_2 (viz 2.2), na ose y procentuálně vyjádřené odpovědi posluchačů. Čerchovaná křivka zachycuje vývoj identifikace stimulů jako /ba/, čárkovaná jako /da/ a tečkovaná jako /ga/.

přechodu. Jak ukazuje obrázek 1 (který zahrnuje navíc ještě přechod mezi /d/ a /g/, vykazující stejné chování), nedochází k postupnému ubývání odpovědi /b/ a komplementárnímu nárůstu počtu odpovědi /d/; namísto toho jsou položky navzdory měnícímu se parametru poměrně dlouho neproblematicky vnímány jako /b/, poté přichází krátké přechodové pásmo, kdy je identifikace nejednoznačná, a následuje opět řada položek vnímaných jako /d/. Stručně řečeno: posluchač nevnímá segmenty řeči objektivně, nýbrž je automaticky řadí do kategorií, jejichž rozlišování má pro něj význam. Tomu je přízpůsobena i jeho rozlišovací schopnost – stimuly nacházející se percepčně na hraně dvou kategorií jsou rozlišovány lépe než stimuly spadající jednoznačně do jedné kategorie. Navzdory dříve předpokládané specifčnosti tohoto fenoménu pro řeč byla kategoriální percepce nalezena i u neřečových stimulů a také u jiných živočišných druhů (Diehl et al., 2004, s. 157–159).

2.1.3 Vliv kontextu na percepci segmentů

Nescházejí doklady o tom, že při percepci posluchač vidí (nebo spíše slyší) dále nežli na jednotlivý segment a že výsledný percepční dojem není na okolí segmentu nezávislý – a to jak na okolí zcela bezprostředním, tak na širším kontextu.

Zaprvé okolí ovlivňuje percepci daného segmentu působením procesů obecně auditivních. Příkladem takového procesu (odehrávajícího se na periférii percepčního aparátu) je krátkodobá sluchová adaptace. Vnímání spektrálních vlastností zvuku je po krátký časový úsek ovlivněno spektrálními vlastnostmi zvuku předcházejícího,

příčemž toto ovlivnění ústí ve zdůraznění rozdílů mezi těmito dvěma stimuly. Důkladem je Summerfieldova studie (Summerfield, Haggard, Foster a Gray, 1984), kde zvuk se zcela plochým spektrem byl posluchači vnímán jako vokál [i] či [a] v případě, že mu předcházel stimul spektrálně komplementární ke schematicky pojatému spektru daného vokálu – to jest vyznačující se nízkou amplitudou ve frekvenčních pásmech, kde se typicky nacházejí formanty onoho vokálu (pro vymezení formantu viz sekci 2.2). Repp (1987a) vyjmenovává řadu dalších studií, které krátkodobou adaptaci dokumentují, a to jak při vnímání neřečových zvuků, tak při vnímání stimulů kombinujících řečové zvuky s neřečovými; a odkazuje dále i na experimenty se zvířecími sluchovými nervy, kde se adaptace taktéž projevila, prokazující tak svou obecně auditivní povahu.³

Že takovéto auditivní mechanismy existují, je nesporné – otázkou je jejich role a důležitost v percepci řeči. Z tohoto hlediska je pozoruhodná studie Holta a Kluendera (2002), jejímž cílem je ukázat, že tyto procesy jsou posluchači využívány ke kompenzaci koartikulace (popsané v předchozím oddílu – 2.1.2). Koartikulace vždy vede k připodobnění sousedících segmentů sobě navzájem, zatímco sluchové procesy uváděné v citované studii působí opačným směrem – to jest zdůrazňují kontrasty, podobně jako výše popsaná krátkodobá adaptace. Podle Holta a Kluendera tohle není šťastná shoda okolností, již se lidé naučili při percepci řeči využívat, nýbrž se v tom odráží skutečnost, že naše percepce je vyvinuta adekvátně pro vypořádávání se s okolním světem. Asimilativní tendence v řeči nejsou ničím ojedinělým – pro veškerý hmotný svět je charakteristické, že pro přechod z jednoho stavu do druhého je třeba jistý čas a energie, pročez dva po sobě následující stavy mohou vykazovat jen omezenou míru odlišnosti. Lidé (stejně jako ostatní živočichové) jsou vybaveni percepčními mechanismy, které jsou této vlastnosti světa uzpůsobeny – a kompenzování koartikulace probíhající již na úrovni obecně auditivní, a nikoliv až specificky řečové, je s tímto předpokladem v naprostém souladu. Další podporu této hypotéze poskytují studie dokládající analogické percepční mechanismy u křepelky japonské (Holt a Kluender, 2002, s. 175).

Výše popsané jevy se týkaly vlivu nejbližšího okolí, avšak vnímání segmentů je

³Citovaná studie je pro tuto práci relevantní především proto, že Repp v ní zkoumá roli krátkodobé adaptace při identifikaci místa artikulace nazálních konsonantů – vrátíme se k ní proto v sekci 2.3.2.

ovlivněno i kontextem výrazně širším. Výmluvným příkladem může být vnější normalizace (Skarnitzl et al., 2016, s. 50), což je mechanismus, kterým se posluchač „naladí“ na konkrétního mluvčího a vyhodnocuje pak vnímané segmenty (například jejich spektrální vlastnosti) s ohledem na specifičnost jeho vokálního traktu. Díky vnější normalizaci můžeme bez obtíží identifikovat tytéž fonémy v řeči statného muže i v řeči dítěte, ačkoliv příslušné segmenty se vlivem značných fyziologických rozdílů akusticky liší. A není to rozhodně záležitost dlouhodobého zvykání si na daný hlas – už několik málo slov může svými vlastnostmi ovlivnit percepci segmentu vyskytujícího se v jejich kontextu (viz klíčová studie Ladefoged, 1957). Dalším významným kontextuálním činitelem může být faktor lexikální. Samuel a Pitt (2003) ve své studii zabývající se kompenzací koartikulace zkoumají, zdali to, že je daný segment součástí jistého sémanticky zatíženého celku, může ovlivňovat percepci tohoto segmentu. Bylo již dostatečně podloženo, že lexikální jednotky hrají roli na vyšší úrovni zpracování řečového signálu (to jest na úrovni „rozhodování“, ovšem nevědomého, do které fonémické kategorie jistý segment zařadíme), avšak Samuel a Pitt hledají jejich vliv na úrovni nižší, v silném smyslu slova *percepční* – a nacházejí jej. Ani sémantickou rovinu řeči tedy nelze při výzkumu percepce segmentů pustit ze zřetele. Jelikož však experiment popsáný v této práci pracoval s pseudoslovy, tedy se stimuly bez lexikálního zatížení, jsou pro nás zajímavější další nálezy citované studie. Výsledky Samuela a Pitta (a dalších jimi citovaných studií) totiž dále naznačují, že percepcí segmentů je ovlivněna jednak pravděpodobností souvškytu daných hlásek (kterou posluchač získává dlouhodobější zkušeností s daným jazykem), jednak percepčním seskupováním, to jest silnější vazbou některých segmentů k sobě navzájem. Vznik těchto vazeb zapříčiňují oba zmíněné faktory (lexikální platnost dané sekvence i vysoká pravděpodobnost jejího výskytu) a jejich existence pak činí takto seskupené segmenty odolnější vůči percepčnímu působení segmentů okolních.

Od vlivu, který má na vnímání segmentů řeči jejich bezprostřední okolí, jsme se tedy přes vliv širšího kontextu a charakteristik vokálního traktu konkrétního mluvčího dostali až k vlivu lexikální roviny a statisticky vyhodnocovaných fonotaktických vlastností daného jazyka. Tato práce se zaměřuje pouze na vliv blízkého segmentálního okolí, ostatní faktory však nebylo možné zcela eliminovat. Přesná podoba experimentu i s ohledem na tuto problematiku je popsána v metodologické sekci

(3.2).

2.2 Percepce místa artikulace okluziv

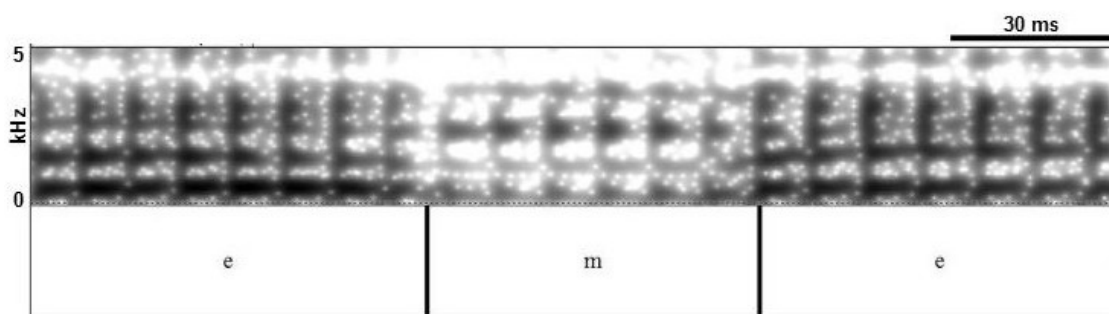
Výzkum percepce nazál byl silně veden výsledky experimentů prováděných s explozivami (viz např. Liberman, Delattre, Cooper a Gerstman, 1954), proto v této sekci shrneme poznatky o nich (případně o okluzivách obecně) a na nazální konsonanty se přímo zaměříme v sekci 2.3.

Výsek akustického signálu, který v segmentálně orientovaném popisu řeči označujeme jako explozivu, obsahuje pouze jediný potenciální zdroj informací o místě artikulace. Skládá se jednak ze závěrové fáze, během níž narůstá tlak v ústní dutině a je buď ticho (u neznělých exploziv), nebo skrze tkáň vokálního traktu uniká základní hlasivkový tón (u exploziv znělých), jednak z exploze, která je důsledkem náhlého uvolnění vytvořeného tlaku. Právě exploze jistou informací o místě artikulace nese: spektrální charakter tohoto velmi krátkého šumu je spoluurčen velikostí a tvarem dutiny, kterou šum prochází (pokud se ovšem za místem závěru nějaká nachází, což neplatí v případě labiálních hlásek) a která působí jako filtr (Johnson, 2012, s. 176). Čistě akusticky vzato by tedy spektrum exploze mělo jednoznačně indikovat místo artikulace. Tento úsek signálu je však velice krátký – po drtivou většinu svého trvání neposkytuje exploziva žádné užitečné informace. Není proto překvapivé, že posluchač při rozpoznávání místa artikulace exploziv využívá především toho, co bezprostředně následuje či předchází (Skarnitzl et al., 2016, s. 54); okolní hlásky⁴ jsou totiž přítomností dané explozivy charakteristickým způsobem ovlivněny (viz sekce 2.1.2).

Klíčovou roli mají v této výzkumné oblasti studie vypracované v 50. letech v laboratořích *Haskins* (zmíněných v sekci 2.1.1 v souvislosti s motorickou teorií percepce). V nich byl pomocí metody *pattern playback*⁵ důkladně zkoumán vliv spektrálních vlastností exploziv a blízkého okolí na vnímání způsobu a místa jejich artikulace (pro shrnutí viz Liberman, 1957). Ukázalo se (Cooper, Delattre, Liberman, Borst a Gerstman, 1952), že i informace nesená explozí samotnou by mohla být percepčně

⁴Výzkumníci si ve svých experimentech volili do sousedství okluziv, pokud je nám známo, téměř výhradně vokály; tato práce nebude výjimkou.

⁵*Pattern playback* je metoda syntetizující zvuk podle ručně nakresleného spektrogramu; umožňuje tedy vytvářet stimuly, které mají sice svou kvalitou do přirozené řeči daleko, zato má však výzkumník spektrum výsledného zvuku zcela pod kontrolou.



Obrázek 2: Spektrogram sekvence [eme] se zřetelnými tranzienty F2.

využitelná, alespoň do jisté míry – identifikace syntetických exploziv jakožto /p/, /t/ nebo /k/ byla zřetelně ovlivněna pozicí exploze na frekvenční ose. K jednoznačnému určení místa artikulace však bylo třeba vyhodnocovat explozi spolu s následujícím vokálem – jakožto invariantní ukazatel se tedy neosvědčila. Výzkumníci proto zaměřili svou pozornost na okolní segmenty.

Stěžejním se v tomto ohledu ukázalo být směřování formantů v části vokálu přiléhající k explozivě. Formanty jsou frekvenční pásma vyznačující se vysokou akustickou energií oproti pásmům ostatním (Skarnitzl et al., 2016, s. 33); z perspektivy filtrové teorie produkce řeči⁶ se jedná se o pásma, která jsou vokálním traktem nejméně tlumena. Poloha těchto pásem koreluje s nastavením nadhrtanových dutin – první formant (F1) odráží zejména polohu jazyka na vertikální ose, druhý formant (F2) na ose horizontální. Zakřivení formantů pozorovatelné ve spektrogramu (například v obou vokálech na obrázku 2) tudíž odráží změnu tohoto nastavení, typicky při přechodu od jedné hlásky ke druhé – proto je tento úsek formantu nazýván tranzient.

Za nejdůležitější je z hlediska identifikace místa artikulace pokládáno směřování formantu druhého (Skarnitzl et al., 2016, s. 54–55), jehož význam byl objeven již v první polovině 50. let (Cooper et al., 1952). Jak se však záhy zjistilo (Delattre, Liberman a Cooper, 1955; Liberman et al., 1954), nejde o směr v absolutním smyslu – zakřivení F2 je charakteristické pro dané místo artikulace jen tehdy, když bereme v úvahu odlišnou pozici F2 charakterizující různé vokály. Invariantním pro dané místo artikulace se v těchto studiích neukázal být přímo jistý tvar tranzientu F2,

⁶Tato teorie nahlíží na produkční aparát jako na systém sestávající ze zdroje (tj. hlasivek), produkujícího poměrně silný a frekvenčně bohatý zvuk, a z proměnlivého filtru (nadhrtanových dutin), který produkovanému zvuku (spolu s radiační impedancí vznikající při přechodu zvuku do vnějšího prostředí) dává jeho výslednou podobu (Fant, 1960).

nýbrž bod, do něhož směřují tranzienty F2 různých vokálů, samozřejmě při zachování stejného místa artikulace sousední okluzivy. Tento bod, zvaný lokus, je bodem pouze virtuálním – tranzienty F2 příslušné hodnoty zpravidla nedosahují a v jednom z experimentů (Delattre et al., 1955) dokonce vedení tranzientů přímo z lokusu zabránilo správné identifikaci příslušné explozivy.

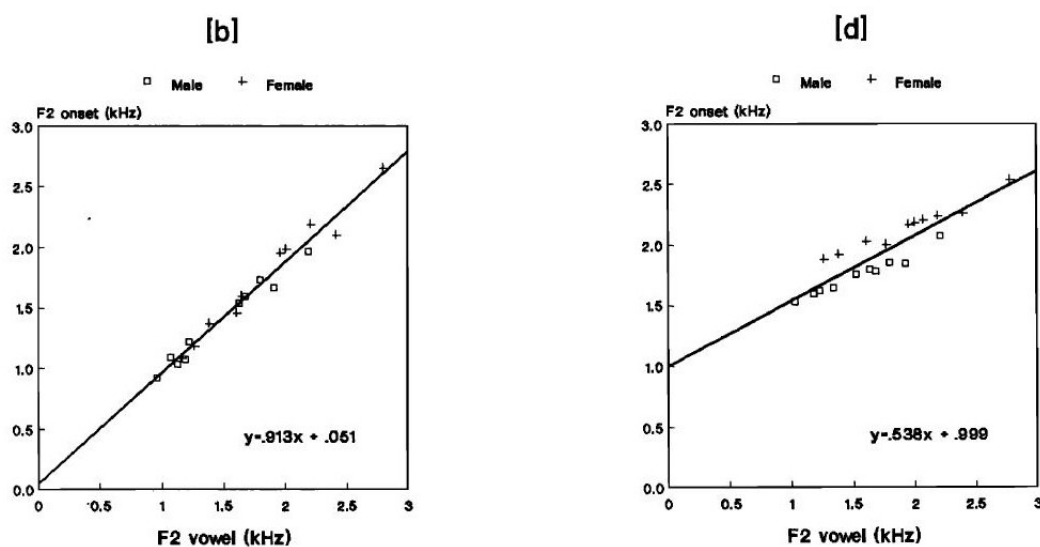
Zde se ovšem výzkum percepce okluziv nezastavil – jednak se i nadále pátralo po jiném invariantním akustickém vodítku, nejlépe přímo přítomném v daném signálu⁷, jednak byla dále prověřována validita teorie lokusu. Potíže se objevily zejména při snaze potvrdit existenci lokusů v přirozené (tj. nikoliv syntetické) řeči (Lehiste a Peterson, 1961; Öhman, 1966); Fant (1969) na základě těchto výsledků koncept lokusu odmítá jako zjednodušující a pro přirozenou řeč nevhodný. Nadějněho „potomka“ našel lokus v teorii tzv. *locus equations*⁸ (Sussman, McCaffrey a Matthews, 1991), v jejímž jádru také leží tranzienty, avšak jednak se obejde bez lokusu jakožto jejich pomyslného průsečíku, jednak se více opírá o statisticky zpracovávaná data pocházející od většího počtu mluvčích. *Locus equations* jsou regresními přímkami proložené body, které získáme, zaneseme-li do grafu hodnoty F2 na samém začátku vokálu spolu s hodnotami F2 v jeho středu, samozřejmě pro různé vokály (viz obrázek 3). Sklon těchto přímek a jejich průsečík s osou *y*, jež zachycuje hodnotu F2 na začátku vokálu, se ukázaly být ve vztahu k místu artikulace slibnými vodítky (Sussman a Shore, 1996)⁹. Také na základě pokusu stanovit polohu tradičně pojatých lokusů prostřednictvím těchto regresních přímek byla existence lokusů (jakožto bodů sice pomyslných, ale přece jen objektivně určitelných z průběhu tranzientů) vážně zpochybněna (Sussman et al., 1991). Tranzienty samotné však ve výzkumu invariantních akustických rysů rozlišujících jednotlivá místa artikulace zůstávají v centru pozornosti; a ani koncept lokusu nepřestal být zmiňován (viz již citovaný Skarnitzl et al., 2016, s. 55).

Navzdory bohatému výzkumu tranzientů okolních hlásek však ani spektrum exploze či vztah jejího spektra ke spektru začátku následující hlásky nebyly smeteny

⁷Pozornost byla věnována zejména tvaru spektra na samém začátku vokálu, jež následuje po explozivě, či vztahu tohoto spektra ke spektru samotné exploze – pro shrnutí viz Raphael (2005, s. 187–188).

⁸Autorka zatím nenašla na zmínku o této koncepci v tuzemské literatuře a nepodařilo se jí nalézt český ekvivalent termínu *locus equations* – je proto v rámci této práce ponechán v angličtině.

⁹Je ale třeba brát na vědomí, že byla testována jejich užitečnost pro strojové rozpoznávání místa artikulace, nikoliv přímo jejich role při percepci.



Obrázek 3: *Locus equations*. Klíčový je odlišný sklon regresní přímky pro [b] a [d] při shodnosti ostatních parametrů. Dále je patrný poměrně malý rozptyl bodů ve vztahu k přímce a také přibližná shoda hodnot u obou pohlaví. Přejato ze studie Sussman et al. (1991).

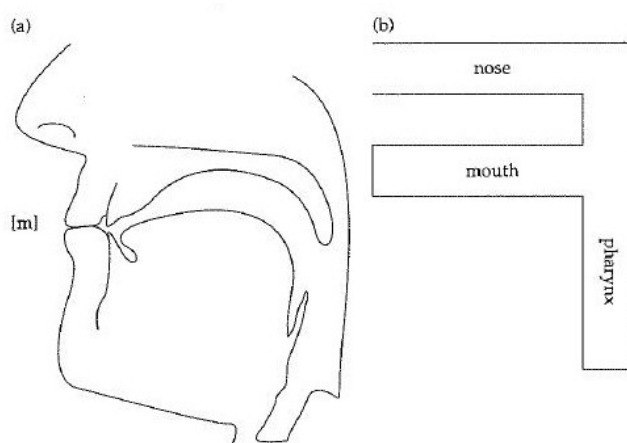
ze stolu. Raphael (2005, s. 189) se na základě důkladné rešerše i svého vlastního výzkumu přiklání k názoru, že relativní percepční důležitost různých akustických korelátů místa artikulace je proměnlivá. Posluchač má k dispozici více funkčních vodítek a patrně mu nic nebrání využít to, které je v dané situaci nejvýraznější (nebo případně jediné dostupné, například vinou nepříznivých akustických podmínek).

2.3 Nazály

2.3.1 Akustické vlastnosti

¹⁰ Tvar vokálního traktu při tvoření nazál lze při hrubém zjednodušení pojmout jako jeden zahnutý tubus (začínající hlasivkami a končící nosem), z něhož vybíhá slepé rameno (ústní dutina, zcela uzavřená jazykem nebo rty); toto schematické pojetí znázorňuje obrázek 4. Jako přibližná délka tubusu se uvádí 21,5 centimetrů a délka slepého ramene se liší v závislosti na místě závěru: nejdelší je u bilabiální nazály a zkracuje se přes alveolární a palatální až po velární (při tvorbě uvulární nazály, která se však v češtině nevyskytuje ani jako foném, ani jako alofon, žádné slepé rameno nevzniká). Pomocí těchto údajů je možné dovést několik charakteristických spektrálních rysů nazálních konsonantů, a především *nazálního brumu*, jak je

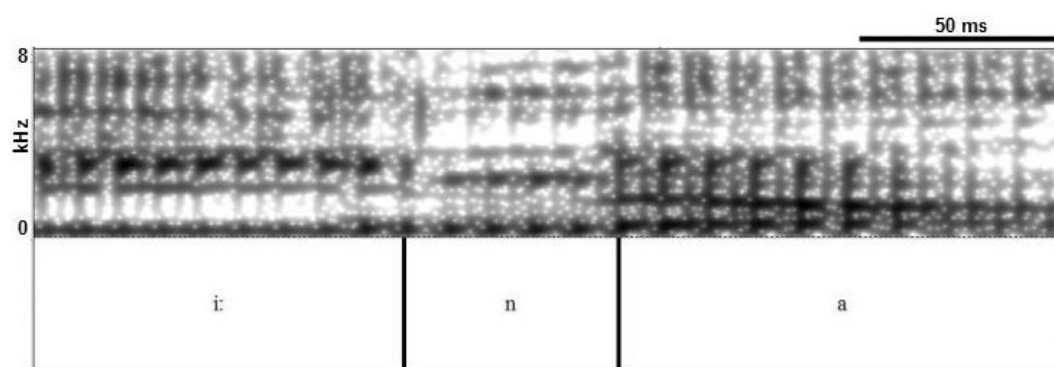
¹⁰Není-li uvedeno jinak, je v této sekci čerpáno z knihy Johnson (2012, kap. 9)



Obrázek 4: (a) Průřez vokálním traktem při produkci hlásky [m]; (b) schematizace jeho tvaru jako soustavy tubusů. Převzato z Johnsona, 2012, s. 190.

nazýváno spektrum onoho frekvenčně značně stabilního úseku mezi závěrem a jeho vypuštěním:

- Formanty nazál jsou tvořeny frekvencemi, jejichž vlnová délka nejlépe „zapadá“ do délky hlavního tubusu (to jest délka tubusu je lichým násobkem čtvrtiny dané vlnové délky – viz Johnson, 2012, kap. 2), takže jsou vlivem stojatého vlnění jejich amplitudy posilovány. Umístění těchto pásem je u všech nazál zhruba shodné.
- Některé frekvence jsou naopak pohlceny slepým ramenem (ústní dutinou), čímž ve spektru vznikají takzvané antiformanty – frekvenční pásma s minimální (v ideálním případě žádnou) akustickou energií. Umístění antiformantů se liší v závislosti na délce onoho slepého ramene – je tedy charakteristické pro jednotlivá místa artikulace.
- Další frekvence jsou pohlcovány menšími slepými rameny tvořenými vedlejšími nosními dutinami; tvar těchto dutin je však výrazně individuální a mění se dokonce i v závislosti na aktuálním stavu mluvčího bez jeho přičinění (například ucpaním při nachlazení). Nehodí se proto k univerzálnímu akustickému popisu.
- Spektrum nazál je co do intenzity celkově slabší než spektrum vokálů, kterým se jinak díky své sonorní povaze podobají, a to kvůli složitější cestě, kterou musí zvuk urazit (je delší, lemují ji vedlejší dutiny a obsahuje značné zúžení na přechodu mezi ústní a nosní dutinou). Také přítomnost antiformantů spek-



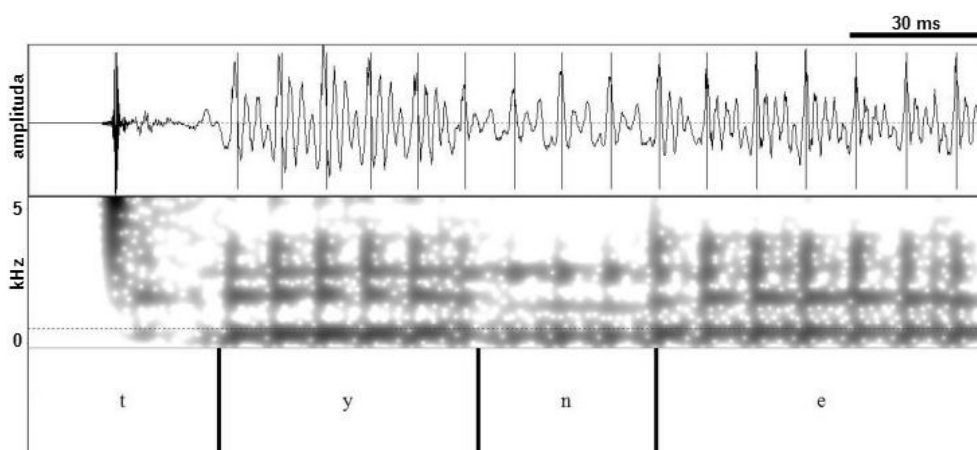
Obrázek 5: Spektrogram sekvence [i:na]. Sytost barvy odpovídá intenzitě příslušné frekvenční složky.

trum oslabuje, a to na všech vyšších frekvencích. Rozdíl v intenzitě spekter nazály a vokálu ilustruje obrázek 5.

K uvedeným rysům je třeba poznamenat dvojí. Zaprvé – případné výpočty poloh formantů a antiformantů nebudou přesně odpovídat hodnotám naměřeným ve skutečně vyslovených nazálech, protože modelování vokálního traktu jako jednoduchého tubusu se slepým ramenem je zjednodušující (neberou se například v úvahu zakřivení a zúžení tubusu). Zadruhé – jediným spektrálním rysem nazálního brumu, který je závislý na místě artikulace, jsou antiformanty, to jest *propady* v akustické energii. Je třeba mít na paměti, že komunikace obvykle neprobíhá ve stabilně tichém prostředí, kde by antiformantová pásma nebyla alespoň částečně vyplněna okolními zvuky – tím se jejich percepční využitelnost potenciálně komplikuje. Ačkoliv v laboratorních podmínkách lze relativní „čistoty“ antiformantů docílit a posluchač je (v souladu s Raphaelovou hypotézou zmíněnou na konci oddílu 2.2) snad i využije jako nejvhodnější vodítka, je možné, že v běžných podmínkách percepčně převažují vodítka méně zranitelná.

2.3.2 Percepce místa artikulace

Nazály se od znělých exploziv artikulacně liší pouze spuštěným měkkým patrem; a akusticky v důsledku toho jednak přítomností nazálních rezonancí během závěrové fáze, jednak o mnoho méně výrazným vypuštěním (unikání vzduchu nosem zabrání vytvoření vysokého tlaku v ústní dutině, a tedy i akusticky výrazné explozi – viz obrázek 6). Co se týče akustických událostí v okolních segmentech, jsou tudíž zjištění popsaná v předchozí sekci aplikovatelná i na nazály, čehož si výzkumníci byli



Obrázek 6: Oscilogram a spektrogram sekvence [tne]; obojí ilustruje rozdíl v akustické výraznosti vypuštění exploziv [t] a nazály [n].

samozřejmě vědomi (např. Liberman et al., 1954; Sussman a Shore, 1996).

Velká artikulační podobnost exploziv a nazál tedy nejprve vedla (v ovzduší rodící se motorické teorie percepce zcela přirozeně, viz 2.1.1) k zaměření se právě na to, co je těmito dvěma třídám hlásek společné – to jest na tranzienty formantů v okolních vokálech. Liberman et al. (1954) dokonce vytvořil pro svůj experiment nazální brum, který měl sloužit pouze jako indikátor nazálního způsobu artikulace a z hlediska místa artikulace měl být zcela neutrální – to mělo umožnit prozkoumání čistě vlivu tranzientů. Spektrální povaha této rezonance byla určena autory na základě jejich percepčního dojmu (zvolili spektrum složené ze tří relativně slabých formantů – okolo 240 Hz, 1020 Hz a 2460 Hz). Následně ji zkombinovali s jedenácti různými průběhy tranzientu F2 v sedmi (taktéž syntetických) vokálech (F1 byl ve všech případech rovný, neboť to podle autorů napomáhalo dojmu nazality). Struktura výsledných slabik byla VC (vokál-konsonant), aby nebyla znevýhodněna nazála [ŋ], která se v angličtině nevyskytuje v přetučě slabiky. Percepční test potvrdil podobný vliv tranzientů na hodnocení jako u exploziv, ač méně silný. To podnítilo další výzkum tranzientů jakožto vodítek pro identifikaci místa artikulace i u této třídy hlásek; a dále tím získala na váze hypotéza silné vazby mezi percepcí a artikulací.

Fáze mezi vytvořením a vypuštěním závěru je však v případě nazál akusticky (a tedy potenciálně také percepčně) bohatší než u exploziv, jak je popsáno v předchozí sekci (2.3.1); záhy proto začala být věnována pozornost také nazálnímu brumu. Řada studií se zabývala jednak relativní důležitostí nazálního brumu a tranzientů v okolních hláskách, jednak rolí spektrální změny odehrávající se v těsné blízkosti

vypuštění závěru. Výsledky některých z těchto studií jsou protichůdné, což však může souviset s rozdíly v použitých metodách.

Malécot (1956) již použil namísto metody *pattern playback* skutečnou řeč nahranou na magnetických páskách, což umožnilo její stříhání a slepování. V jeho experimentech měli proto respondenti za úkol identifikovat jednak samostatné nazální brumy (jak nahrané izolovaně, tak vystřížené ze slabiky), jednak slabiky lepené ze dvou typů zdrojů – z hlásek [m, n, ŋ, æ] nahraných v izolaci a z rozstříhaných slabik [mæ, næ, ŋæ, æm, æn, æŋ]. Vznikly tedy mimo jiné stimuly obsahující teoreticky protichůdné akustické informace – například [m] připojené k vokálu vystříženému ze slabiky [æŋ]. Výsledky percepčního testu, v němž měli respondenti za úkol vždy identifikovat danou nazálu, hovořily pro percepční převahu tranzientů nad nazálním brumem, ačkoliv ani ten nebyl při identifikaci zcela bezvýznamný. Doplňující experiment, v němž Malécot lepil k nazálním brumům vokály pocházející ze sousedství hlásek [b, d, g] pak ještě silněji podpořil hypotézu o vazbě artikulace a percepce.

Recasens (1983), zkoumající identifikaci nevypuštěných nazálních konsonantů ve finální pozici katalánskými posluchači, pracoval opět se syntetickými stimuly (vytvořenými však již jinou metodou než *pattern playback*). Na rozdíl od Libermana nepoužil jen jeden „neutrální“ nazální brum, nýbrž na základě výsledků akustických měření provedených na nahrávkách skutečně vyslovených slabik vytvořil nazální brumy pro jednotlivé nazály s odpovídajícími spektrálními vlastnostmi. Ty pak, podobně jako předchozí studie, kombinoval s různými tranzienty F2 a F3. Na základě výsledků percepčního testu se taktéž přiklonil k převaze tranzientů nad nazálním brumem při identifikaci místa artikulace, přičemž však uznával významnou roli brumu zejména v případě, že tranzienty sousedního vokálu (v této studii [a]) jsou nejednoznačné.

Následoval výzkum Kurowské a Blumsteinové (1984), které nepracovaly ani se syntetickou řečí, ani se stimuly lepenými, nýbrž pouze s výřezy z přirozených nahrávek. Důvodem byla snaha zkoumat skutečnou řeč a vyvarovat se při tom případného vlivu nepřirozeného akustického skoku přítomného v lepených stimulech. V jejich experimentu byly posluchačům prezentovány úseky slabik sestávajících z nazál [m, n] a vokálů [i, e, a, o, u], lišící se jednak délkou (kratší obsahovaly pouze šest period základní frekvence, delší byly tvořeny vždy celou nazálou či vokálem), jednak místem,

z něhož byly vyříznuty (kratší stimuly pocházely zaprvé z úseku těsně před vypuštěním závěru, zadruhé z úseku těsně po něm a zatřetí z úseku obsahujícího poslední tři periody nazály a první tři vokálu). Výsledky percepčního testu ukázaly, že nazální brum je z hlediska místa artikulace zhruba stejně informativní jako tranzienty následujícího vokálu, což je v rozporu s výsledky Malécota (1956) i Recasense (1983). Je však třeba si všimnout, že v důsledku odlišné metody se výzkumné otázky zmíněných studií poněkud liší – zatímco Kurowski a Blumstein zkoumaly, zdali nazální brum a tranzienty obsahují zhruba stejně efektivní vodítka pro identifikaci místa artikulace, Malécot a Recascens se zaměřovali na otázku, které z těchto vodítek zvítězí při percepci, jsou-li obě přítomna a ve vzájemném konfliktu. Ani jeden z nich všal nezpochybňuje, že i nazální brum může být pro percepci místa artikulace významný (zejména v případě sekvencí VC či nejednoznačnosti tranzientu).

Důležitým pro směřování dalšího výzkumu byl pak také ten nálezný citované studie (Kurowski a Blumstein, 1984), že nejlépe byly identifikovány stimuly obsahující jak část nazály, tak část následujícího vokálu. Posluchač tedy patrně nevolí pouze jedno z těchto vodítek, nýbrž vnímá obě a na nějaké úrovni zpracování dané slabiky je integruje. Autorky studie se přiklánějí k názoru, že tato integrace probíhá již na velmi brzkém stupni zpracování – tedy že posluchač již *vnímá* nazální brum a tranzienty následující hlásky jako jedinou vlastnost slabiky, a nikoliv že percipuje tato vodítka odděleně a následně je na nějaké vyšší úrovni zkombinuje. K hypotéze brzké integrace se ovšem autorky, jak samy uvádějí, přiklánějí pouze pro její větší jednoduchost – samotné výsledky jejich experimentu neposkytují pro toto rozhodnutí dostatečný podklad.

Repp (1986) navázal na studii Kurowské a Blumsteinové a provedl řadu experimentů s různě upravenými nahrávkami skutečné řeči (pocházejícími od šesti mluvčích, nikoliv pouze od jednoho jako u Kurowské a Blumsteinové), jimiž výsledky těchto autorek potvrdil – i v jeho testech přítomnost nazálního brumu i vokálu s tranzienty výrazně napomáhala identifikaci místa artikulace. Pomocí vkládání krátkých úseků šumu na různá místa hodnocených slabik a dále pomocí slabik složených z nazálních brumů a vokálů zbavených všech dynamických prvků také získal výsledky potvrzující hypotézu poměrně brzké integrace; jistě nejsou tato dvě vodítka zpracována jako dvě separátní informace a následně na nějaké abstraktní ro-

vině vyhodnocována. Zároveň ale Repp ze schopnosti posluchačů integrovat tato dvě vodítka i tehdy, když klíčový úsek slabiky byl překryt šumem, vyvozuje, že integrace se pravděpodobně neodehrává přímo na úrovni nejnižší, čistě auditivní.

Další výzkum posunul tuto otázku ještě dále a ke konci 80. let se do popředí dostává hypotéza, že klíčovým vodítkem (alespoň v případě slabik se strukturou CV; k VC slabikám se vrátíme níže) je přímo spektrální změna odehrávající se na hranici nazály a sousedního vokálu (Kurowski a Blumstein, 1987; Repp, 1987a, 1987b), tedy ani nazální brum, ani tranzienty v sousedním vokálu, ani tato dvě vodítka integrovaná dohromady (ať už na jakémkoliv stupni zpracování řečového signálu). Na tomto místě je vhodné připomenout Reppův (1987a) pokus propojit otázku percepce místa artikulace nazál s fenoménem krátkodobé sluchové adaptace (2.1.3). Hypotézu, že by právě tento fenomén byl klíčový, Reppovy výsledky ve svém celku nepotvrzují; ukazují však alespoň, že posluchači při percepci významně těžší z těsného sousedství nazály a vokálu, ač patrně jinak než pomocí krátkodobé adaptace; právě to vede Reppa k zaměření pozornosti přímo na spektrální změnu na přechodu segmentů. Kurowski a Blumstein (1987) podníceny tímto vývojem pátraly v CV slabikách vyslovených několika mluvčími po měřitelném akustickém parametru v těsném okolí vypuštění nazály, který by se osvědčil při automatickém rozlišování [m] a [n]. Slibným se jim jeví pozorování, že v případě [m] dochází při přechodu od nazály k vokálu k výraznější změně intenzity ve vyšším frekvenčním pásmu (11–14 Barků), kdežto v případě [n] v pásmu nižším (5–7 Barků). Částečně se jim tento parametr při klasifikaci osvědčil, avšak nefungoval stejně dobře pro všechny samohláskové kontexty (viz níže v tomto oddílu).

Co se týče důležitosti spektrální změny v místě vypuštění nazály, je však třeba opatrnost zejména v případě nazál ve finální pozici – výsledky percepčních testů pro slabiky se strukturou VC se podle Reppa (1987b) od výsledků CV slabik nepadatelně liší.¹¹ Krátký úsek přechodu mezi vokálem a nazálou při percepci VC slabik nehrál zásadní roli (tento přechod ostatně bývá v signálu poměrně málo zřetelný); posluchači se zdáli spoléhat mnohem spíše na nazální brum a tranzienty jakožto dvě nezávislá, kombinovatelná vodítka, což pro CV slabiky naplatilo (viz

¹¹Jistého rozdílu si všiml už Malécot (1956) – důležitost nazálního brumu pro percepci je ve VC slabikách podle jeho výsledků vyšší (což může souviset s delším trváním nazál ve finální pozici, které na svých vzorcích zpozoroval).

Repp, 1986 a zejm. Repp, 1987a). Repp tedy vyslovuje domněnku, že v těchto dvou situacích používáme odlišné percepční mechanismy; to by bylo v souladu s již zmiňovaným tvrzením Raphaela (2005), že posluchač si z více dostupných vodítek volí to v daném kontextu nejvhodnější.

Faktorem, který nemůžeme při zkoumání percepce nazál opomenout, je kvalita sousedního vokálu. Popsali jsme již její funkci při stanovování tradičně pojatého lokusu, avšak tam její úloha patrně nekončí. Studie zahrnující více vokalických kontextů narážely na nezanedbatelný vliv této proměnné i tehdy, když braly vokalickou specifičnost tranzientů v potaz. Liberman et al. (1954) uvádí, že domnělý „labiální“ tranzient F2 vedl k percepci hlásky [m] nápadně méně často u vokálů [i, e] než u vokálů [ɛ, a, ɔ, o, u]; výsledky Nakaty (1959) naznačují, že pozice lokusu charakterizujícího danou nazálu se se změnou vokálu trochu posouvá (což ovšem oslabuje koncepci lokusu jako takovou); v experimentu Kurowské a Blumsteinové (1984) byla úspěšnost při rozpoznávání místa artikulace (jak ze samotného nazálního brumu a samotných tranzientů, tak z krátkých výseků obsahujících vypuštění závěru) různým způsobem ovlivněna kvalitou vokálu ve zdrojové slabice; Reppovi se v jedné studii (1986) u pěti různých typů stimulů ukázala horší identifikovatelnost místa artikulace ve slabikách s [i] než ve slabikách s [a, u] a ve studii navazující (1987a) šly výsledky percepce slabik s [i] spíše proti Reppově hypotéze krátkodobé adaptace, zatímco percepce slabik s [a, u] s ní byla v souladu; Kurowské a Blumsteinové (1987) při pokusech s automatickou klasifikací vyšel jistý akustický korelát místa artikulace jako poměrně úspěšný parametr – avšak pro slabiky s [i], popř. s [i, e] (v případě stimulů formy [sCV], tedy [s]+konsonant+vokál), příliš dobře nefungoval¹². Percepce nazál v kontextu předních a patrně spíše vysokých vokálů tedy zjevně vyžaduje zvláštní pozornost.

Jak dokládá celá tato sekce, kandidátů na invariantní akustický korelát místa artikulace se postupně objevilo více a jejich relativní důležitost pro percepci byla hodnocena různě. Situaci značně komplikuje citlivost zkoumaných parametrů jak na vokalický kontext, tak na pozici nazály ve slabice; navíc je vztažení výsledků dosavadních studií k sobě navzájem v některých případech ztíženo výraznou od-

¹²Kurowski, Blumstein v této studii upozorňují na další výzkumy akustických vlastností slabik s nazálami, v nichž se slabiky obsahující [i] chovaly specificky.

lišností použitých metod. Publikace sumarizačního či příručkového charakteru se ovšem často drží spíše tradičnějšího výkladu, který zdůrazňuje (či uvádí výhradně) úlohu tranzientů sousedního vokálu – viz například Clark a Yallop (1995, s. 317); Raphael (2005); Mitterer a Cutler (2006, s. 774), kde je ovšem alespoň zohledněn rozdíl mezi CV a VC slabikami; Skarnitzl et al. (2016, s. 57).

3 Metoda

3.1 Zvukový materiál

S laskavým svolením KKY se v této studii pracovalo s nahrávkami obsaženými v korpusu vytvořeném na tomto pracovišti pro účely syntézy řeči. Využíván byl mluvčí, jehož syntetická řeč byla v době vytváření studie považována za nejpracovanější a u nějž byl zároveň původně zpozorován jev zkoumaný v této práci¹³. Mluvčí byl nahráván v bezdozvukové komoře na profesionální studiový mikrofon a segmenty získané automaticky z jeho řeči byly uloženy se vzorkovací frekvencí 16000 Hz a v komprimované podobě¹⁴. Při vytváření této práce byla využívána stále táž verze korpusu (kód /artic_images/spkr_AJ.rev661.img).

Pro vytvoření percepčního testu byly využity výhradně nahrávky pocházející přímo z korpusu, nikoliv řeč syntetizovaná. Ačkoliv totiž i zde pracujeme s konkatencí, na rozdíl od původní syntézy považujeme za základní jednotky tradiční segmenty odpovídající hláskám, nikoliv difony. Tento přístup byl zvolen jednak proto, že lépe umožní vztáhnout výsledky této studie ke studiím předchozím, jednak proto, že tato práce je zaměřena na zkoumání vlivu hláskového okolí na percepci daného segmentu a diskontinuita uvnitř segmentu samotného, ke které plzeňský způsob segmentace vede, by mohla být faktorem ovlivňujícím percepci, jenž by s hláskovým okolím souvisel jen nepřímo.

3.2 Tvorba stimulů

Percepční test obsahoval tři typy stimulů:

1. Nejdůležitější skupinou byly stimuly lepené ze dvou zdrojových slov či sousloví, jež obě obsahovala intervokalické „n“. Zdrojové promluvy byly manuálně rozděleny mezi nazálou a jí předcházejícím vokálem (při segmentaci byla dodržována pravidla popsaná v publikaci Machač a Skarnitzl, 2009) a ze získaných úseků vznikla prostým slepením pseudoslova – příkladem může být slovo „klarinechal“, utvořené slepením úseků „klari“ ze slova „klarinetu“ a „nechal“ ze slova „ponechal“. V testu pak

¹³O tom, že nejde o zcela ojedinělý případ týkající se pouze jednoho mluvčího, však svědčí fakt, že u jedné další mluvčí z téhož korpusu je pozorovatelná táž tendence (opět ve slově „mlýneček“). Viz nahrávka v Příloze 2.

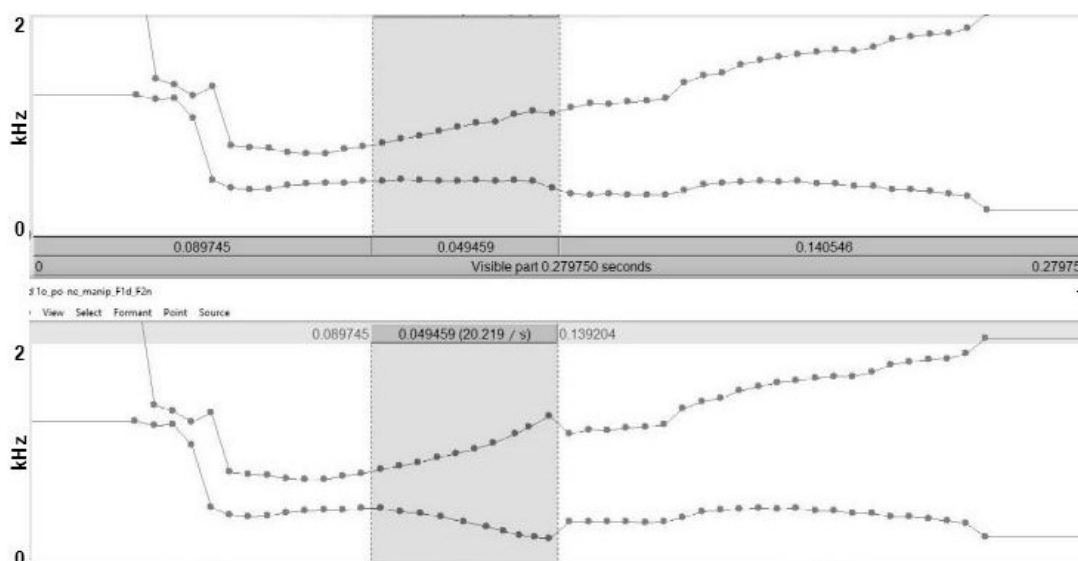
¹⁴Metoda ADPCM (5bit)

byly použity pouze dvojslabičné sekvence vyříznuté z těchto pseudoslov (v případě uvedeného slova by šlo o sekvenci „rinech“). Hlavním důvodem pro použití výřezů byla snaha o potlačení možného vlivu sémantiky na percepci. Ovlivnění percepce významem by mohlo hrozit i přesto, že se přísně vzato jedná o pseudoslovo – i ve výše uvedeném příkladu jsou obě jeho složky snadno identifikovatelné. Pseudoslova byla vybrána tak, aby mezi nimi byly jak neproblematické případy, kdy segment [n] zní i po slepení jako [n] (aby bylo zjevné, že ke zkoumanému jevu nedochází vinou rozříznutí a slepení jako takového, nýbrž že klíčovou úlohu hraje to, k čemu byl sledovaný úsek připojen), tak případy, které by mohly ilustrovat jev popsáný výše (1.1), kdy po slepení vzniká spíše dojem hlásky „m“ (že tomu tak alespoň v některých případech je, bylo ověřeno v pilotní fázi testování). Zdrojové promluvy byly k vytváření stimulů využívány opakovaně – například slabiky obsahující nazálu pocházející z pouhých tří slov („ponechal“, „Janečka“ a „fingovaná“), aby bylo možné doložit, že jedna a táž slabika (nikoliv pouze dvě slabiky stejného fonémického složení) může být vlivem slabiky předchozí percipována různě.

2. Druhou skupinou stimulů byly výřezy ze zdrojových promluv, svou délkou a strukturou přibližně odpovídající výřezům z pseudoslov (například ze slova „ponechal“ byl použit úsek „pone“). Jde o promluvy využití jednak pro lepení pseudoslov, jednak pro tvorbu manipulovaných položek, které jsou popsány v následujícím bodě. Právě kvůli tomuto typu stimulů bylo nezbytné použít v testu pouze krátké výřezy, neboť zde by byl tlak sémantiky na percepci neoddiskutovatelný. Test neobsahoval všechny zdrojové promluvy použité pro lepení pseudoslov, a to v zájmu jeho rozumného rozsahu. Vynechány však byly pouze takové zdroje, které se neúčastnily tvorby pseudoslov, jež se na základě poslechu autorky a výsledků pilotních testů zdála být potenciálním dokladem sledovaného jevu.

Do této skupiny spadají také výřezy ze tří zdrojových promluv, v nichž mluvčí zcela záměrně produkoval intervokální „m“ – tyto stimuly byly zařazeny proto, aby respondenti byli konfrontováni také s přirozeným „m“ tohoto mluvčího, a ne pouze s podobami této hlásky vzniknuvšími následkem zde sledovaného jevu. Z tohoto důvodu byly další dva takové stimuly zařazeny také do zácvičku (jehož podoba bude blíže popsána níže).

3. Třetí, nepočatnou skupinou byly výřezy z promluv (čtyř zdrojových a jedné



Obrázek 7: Trajektorie F1 a F2 v sekvenci „pone“ před manipulací (nahore) a po ní (dole). Úsek, kde bylo s formanty manipulováno, je vyznačen šedě.

lepené), v nichž bylo manipulováno formanty vokálu předcházejícího nazále (zpravidla formantem prvním a druhým, v jednom případě také třetím). Cílem bylo pokusit se pouhou manipulací tohoto vokálu způsobit zkoumaný jev, pozorovaný dosud jen u zvuků vzniklých lepením – to jest docílit toho, aby nazála v manipulovaném slově byla vnímána jako „m“ či aby její status byl přinejmenším nejednoznačný. Pokud by se toto alespoň částečně podařilo, byla by tím otevřena možnost využití tohoto nástroje pro další experimenty – zatímco při slepování úseků slov se vždy nutně mění celá řada parametrů, které výzkumník nemá pod kontrolou, manipulace s formanty umožňují měnit okolí nazály poměrně dobře kontrolovatelným způsobem a sledovat vliv různých drobných, na sobě nezávislých změn. Manipulace nebyly vedeny žádnou hypotézou stanovující, jakým způsobem či v jakém rozsahu by se měly formanty upravit – použité položky byly vybrány z velkého množství rozličných pokusů na základě subjektivního soudu autorky korigovaného třemi obětavými posluchači. Tuto část experimentu je tudíž třeba považovat za ve velmi silném smyslu pokusnou. Příklad provedených manipulací je zachycen na obrázku 7.

Manipulace byly provedeny pomocí softwaru Praat (Boersma a Weenink, 2015) následujícím postupem: Zvuk byl převzorkován z původní vzorkovací frekvence 16000 Hz na frekvenci 11000 Hz; metodou LPC¹⁵ z něj bylo extrahováno pět formantů

¹⁵Přesné parametry: LPC (covariance); řád: 10; délka segmentačního okénka: 0,025 s; časový krok: 0,005 s; preemfáze: 50 Hz. Toto nastavení bylo použito pro všechny položky.

a excitační signál; formanty byly manuálně upraveny a poté jimi byl excitační signál filtrován. Parametry metody byly nastaveny tak, aby při co nejuspokojivější kvalitě výsledného zvuku metoda extrahovala formanty způsobem umožňujícím snadnou manipulaci s prvními třemi z nich; zvolené nastavení je tedy kompromisem mezi těmito dvěma požadavky. Výsledný zvuk se ovšem od výchozího lišil nejen upravenými formanty – použitá metoda měla nevyhnutelné důsledky pro jeho celkový charakter (změna spektrálního složení byla jasně patrná jak vizuálně, tak auditivně). Jelikož manipulované položky by se z popsaných důvodů svou kvalitou výrazně lišily od ostatních, byl i na všechny další stimuly aplikován postup popsaný v tomto odstavci, pouze s tím rozdílem, že s formanty nebylo nijak manipulováno. Tím se docílilo shodného charakteru všech položek.

Metoda popsaná v předchozím odstavci si vyžádala další drobné úpravy, provedené v programu Audacity (Audacity Team, 2015). U několika položek měla za následek značně nepřírozenou a nepříjemnou podobu některých frikativ a afrikát; a jelikož tyto hlásky nejsou pro prováděný experiment nijak důležité, byly v zájmu větší přirozenosti položek nahrazeny příslušnými segmenty z původních nahrávek. Nakonec byla upravena hlasitost všech položek tak, aby byly percepčně všechny na téže úrovni.

V sekci 2.1.3 je zmíněna řada faktorů, které percepci segmentů ovlivňují a z nichž představovaný experiment cílí pouze na jeden, totiž na nejbližší segmentální okolí. Vliv vnější normalizace byl vzhledem k využití jediného mluvčího konstantní, vliv lexika byl potlačen použitím pseudoslov. Na pravděpodobnost výskytu použitých sekvencí nebyl brán zřetel, jelikož hlavním kritériem při výběru položek byl předpokládán nejednoznačný percepční status obsažené nazály a počet vhodných kandidátů nebyl tak velký, aby bylo možné důsledně tento fonotaktický faktor ošetřit.

3.3 Percepční test

Percepční test obsahoval celkem 35 stimulů s následujícím zastoupením skupin popsaných v předchozí sekci: 15 stimulů lepených; 15 výřezů ze zdrojových promluv (nelepených a nemanipulovaných); 5 stimulů manipulovaných (z nichž jeden pocházel z promluvy lepené, čtyři ze zdrojových). Každý stimul zazněl dvakrát, celkový počet položek byl tedy 70. Pořadí položek bylo pro každého respondenta náhodné

(s jediným omezením, totiž že totožné položky nesměly následovat hned za sebou). Byl zvolen typ úlohy 2AFC (*two-alternative forced choice*): U každé položky byla mluvčímu nabídnuta dvě pseudoslova, z nichž měl zvolit to, které odpovídalo slyšenému stimulu¹⁶. Možnosti se lišily pouze tím, že v jedné bylo mezi vokály „n“ a ve druhé „m“. V případě nejistoty měl respondent možnost pustit si daný stimul znovu, nejvýše však třikrát. Každému stimulu (i pokud byl opakován) předcházelo krátké pípnutí. Test byl rozdělen na čtyři bloky (první tři obsahovaly po 20 položkách, poslední 10), mezi nimiž si respondent mohl (a byl k tomu v instrukcích veden) udělat přestávku pro obnovení pozornosti. Celkové trvání testu se však díky krátkosti položek pohybovalo kolem 6 minut včetně zácviků, ochabnutí pozornosti vinou únavy tedy příliš nehrozilo.

Zácvik byl koncipován tak, aby jednak umožnil respondentům zvyknout si na syntetický hlas (navíc poškozený procedurou nutnou kvůli manipulačním formantům, viz oddíl 3.2), jednak je seznámil s podobou testu. Obsahoval proto dvě výrazně delší položky (celé věty), jednu položku trojslabičnou a čtyři položky již zcela odpovídající položkám v samotném testu; pořadí položek v zácviku bylo ovšem pevné. Mluvíci byli na specifickou podobu zácviků v instrukcích upozorněni (viz poznámka 16).

Test včetně zácviků byl vytvořen pomocí nástroje ExperimentMFC v softwaru Praat (Boersma a Weenink, 2015). Před zahájením testování byla provedena pilotáž, jíž se zúčastnilo celkem pět respondentů; na základě jejich komentářů byly provedeny drobné úpravy ve formulaci zadání.

3.4 Respondenti

Experimentu se zúčastnilo celkem 32 posluchačů (18 mužů a 14 žen) v průměrném věku 25,8 let (od 17 do 40), bez diagnostikovaných či dotyčným samým pozorovaných sluchových vad. Test byl (až na dvě výjimky) zadáván na dálku, přičemž

¹⁶Instrukce vypadaly doslova takto: „V následujícím testu uslyšíte řadu nesmyslných slov. Vaším úkolem bude vždy vybrat ze dvou možností tu, která slyšenému slovu odpovídá. V případě potřeby si můžete položku přehrát znovu (nejvýše třikrát). Je nutné vždy jednu z možností vybrat, i pokud si nejste jisti. Test obsahuje celkem 70 krátkých položek; po každé dvacáté následuje přestávka, kterou prosím využijte ke krátkému odpočinku. Postup testu můžete sledovat v levém horním rohu okna. V testu je použita syntetická řeč, zácvik proto obsahuje dvě delší položky, které Vám umožní se s mluvčím seznámit. Během zácviků si také nastavte hlasitost na příjemnou hladinu. Nemáte-li nyní otázky, kliknutím zahajete zácvik.“ Po zácviků, těsně před začátkem samotného testu, byl respondent ještě upozorněn, že počty položek obsahujících „n“ a položek obsahujících „m“ nemusejí být vyrovnány, aby případná snaha o vyváženost typů odpovědí neovlivňovala hodnocení.

respondentům byla vysvětlena nezbytnost tichého, klidného prostředí a především použití sluchátek. Žádný z respondentů nehlásil potíže s průběhem testu či jakékoliv nestandardní události.

3.5 Zpracování dat

Výsledky percepčního testu byly zpracovány pomocí programu R (R Development Core Team, 2016). Odpovědi respondentů byly převedeny na číselnou škálu ($n = -1$, $m = 1$) a bylo vypočteno průměrné hodnocení každé položky každým posluchačem (neboť všechny položky se v testu vyskytly dvakrát). Následně byly pomocí metody bootstrap¹⁷ získány průměry (a jejich konfidenční intervaly) pro jednotlivé položky rozdělené do tří skupin – zdrojové stimuly, lepené stimuly a manipulované stimuly. Zvolená $\alpha = 0,05$ byla vždy upravena Bonferroniho korekcí, čímž byl zohledněn počet položek v jednotlivých skupinách. Výsledky budou prezentovány formou grafů.

V zájmu statistického potvrzení závěrů plynoucích z porovnání grafů byl u vybraných dvojic položek (sestavajících vždy ze stimulu lepeného či manipulovaného a k němu příslušícího zdroje) následně statisticky ověřen významný rozdíl v jejich hodnocení. Jelikož se jednalo vždy o dva závislé výběry, byl použit párový Wilcoxonův test (tento test je neparametrický, rozdělení dat tedy nehrálo roli).

Výsek z tabulky zobrazující zpracovávaná data a klíčové úseky z použitého skriptu jsou obsahem Přílohy 3.

¹⁷Metoda bootstrap vytváří z výběrového souboru získaného sběrem dat další výběrové soubory pomocí tzv. náhodného výběru s opakováním – některé hodnoty původního výběrového souboru opakuje a jiné vynechává. Pomocí těchto uměle vytvořených výběrových souborů lze pak lépe simulovat vlastnosti souboru základního (v tomto případě průměrné hodnocení daného stimulu) i v případě nenormálního rozdělení dat.

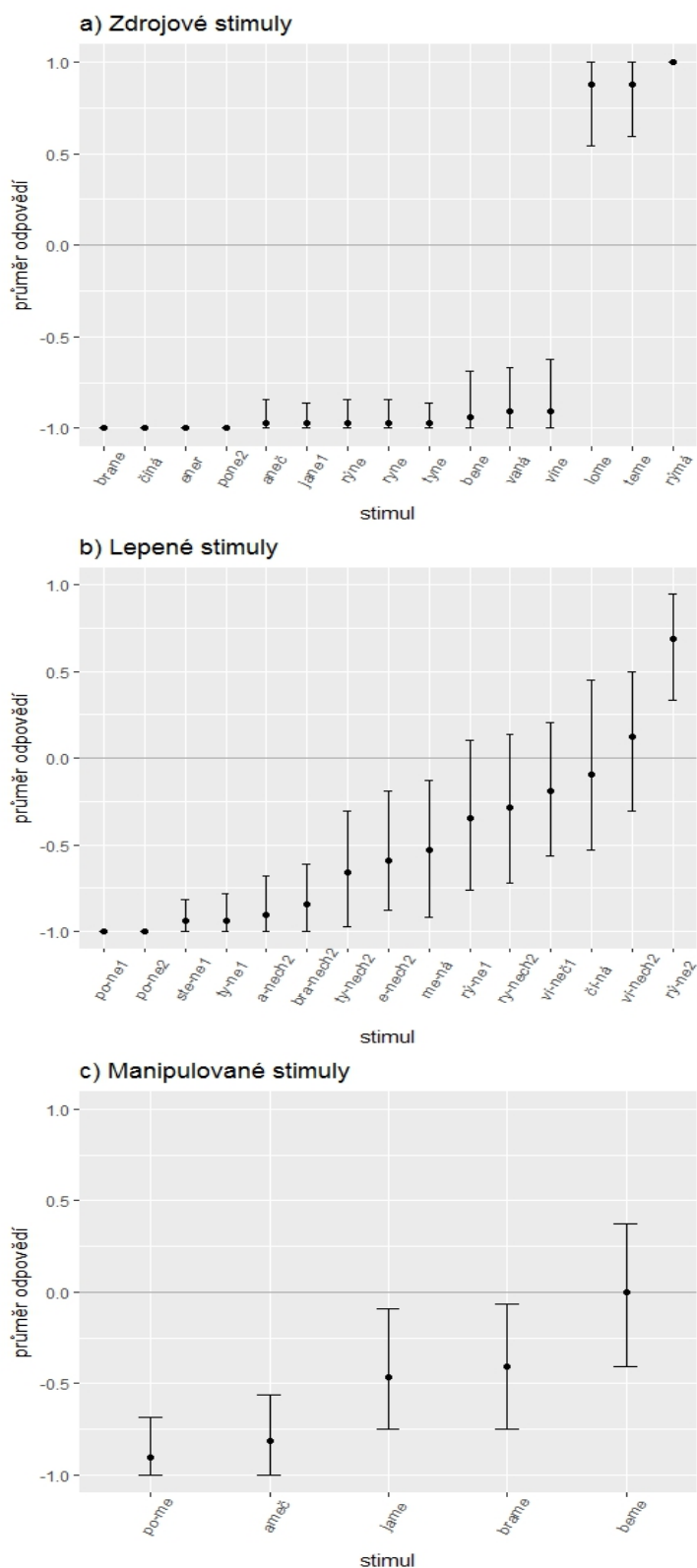
4 Výsledky

Grafy 8a, 8b a 8c zachycují průměrné hodnocení jednotlivých položek všemi posluchači. Hodnota 1 na ose y odpovídá zvolení varianty obsahující /m/, hodnota -1 zvolení varianty s /n/. Hodnota 0, jež je v grafech zdůrazněna, tedy dělí graf na dvě části – v položkách, jejichž průměr spadá do horní části, bylo slyšeno spíše [m], a v těch, jejichž průměr leží v části spodní, spíše [n]. Je však třeba brát v úvahu kromě samotného průměru také konfidenční interval – při zvolené $\alpha = 0,05$ může skutečná střední hodnota ležet kdekoliv v jeho rozsahu. Kdykoliv tedy konfidenční interval protíná nulovou hodnotu, nelze než prohlásit hodnocení dané položky za nejednoznačné.

Graf 8a zobrazuje zdrojové (tedy nelepené a nemanipulované) stimuly. Hodnocení všech dvanácti položek obsahujících [n] i tří kontrolních položek obsahujících [m] je zcela jednoznačné; shoda respondentů však ani u těchto stimulů není vždy stoprocentní (takového výsledku bylo dosaženo pouze v případě pseudoslov „brane“, „číná“, „ener“, „pone“ a „rýmá“)¹⁸.

Graf 8b zahrnuje všechny lepené stimuly (celkem 15 položek). Číslo za názvem položky udává, ze kterého zdrojového slova pochází slabika „ne“ (případně „nech“ či „neč“) – 1 značí slovo „Janečka“, 2 slovo „ponechal“ (tytéž indexy nesou i výřezy z těchto dvou slov v grafu 8a zachycujícím zdrojové stimuly). Stimul „čí-ná“ žádný index nenesou, neboť jeho zdroj „vaná“ je jedinečný. U devíti položek byla respondenty významně častěji volena varianta obsahující [n] (u dvou z toho bezvýjimečně). V případě pseudoslov „rýne“, „rynech“, „víneč“ a „číná“ spadá sice nalezený průměr do spodní části grafu (a tedy ukazuje spíše na [n]), ale konfidenční intervaly protínají hraniční linii – jejich hodnocení tedy musí být označeno za nejednoznačné, stejně jako hodnocení položky „vínech“, jejíž průměr se nachází nad hraniční linií, ale konfidenční interval ji opět protíná. Pouze položka „rýne“ byla jednoznačně častěji hodnocena jako obsahující [m]. Tyto výsledky potvrzují, že segmentální okolí má jistý vliv na percepci místa artikulace nazál – pouhým nahrazením předchozí slabiky slabikou jinou (pocházející však ze sousedství téhož fonému) se hodnocení nazály stalo nejednoznačným.

¹⁸Odchytky od většinového hodnocení mohou souviset i s nepříliš dobrou kvalitou nahrávek zaviněnou filtrací; to je však jen jeden z možných vlivů.



Obrázek 8: Grafy zachycující průměry hodnocení všech položek percepčního testu, rozdělených na a) lepené stimuly, b) manipulované stimuly a c) zdrojové stimuly, spolu s konfidenčními intervaly ($\alpha = 0,05$). Volbě odpovědi s [n] v percepčním testu odpovídá hodnota -1 na ose y , volbě odpovědi s [m] hodnota 1 .

Graf 8c zachycuje hodnocení položek, s jejichž formanty bylo manipulováno. Pouze jedna, pseudoslovo „bene“ s F1 a F2 uměle zahnutými směrem dolů, byla hodnocena nejednoznačně (průměrná hodnota se dokonce nachází přímo na hraniční linii), u ostatních byla významně častěji volena varianta s [n].

Wilcoxonův test byl proveden pro ty položky, jejichž konfidenční interval protíná hraniční hodnotu nebo celý spadá do horní části grafu (tedy „rý-ne1“, „ry-nech2“, „ví-neč1“, „čí-ná“, „ví-nech2“, „rý-ne2“ z lepených stimulů a „beme“ z manipulovaných). Hodnocení dané položky bylo vždy porovnáno s hodnocením slova, z něhož pocházela slabika [ne] – byla tudíž testována přímo významnost posunu hodnocení konkrétní slabiky při změně slabiky předchozí. Ve všech sedmi případech byla výsledkem testu hodnota $p < 0,001$, tudíž přijímáme alternativní hypotézu: Hodnocení jmenovaných položek se liší od hodnocení k nim náležejících zdrojových slov na hladině významnosti $\alpha = 0,05$. Tento výsledek lze ostatně vyčíst i ze samotných grafů – konfidenční intervaly příslušných stimulů se nepřekrývají.

5 Diskuse

5.1 Interpretace výsledků

Hlavním cílem této práce bylo doložit vliv segmentálního okolí na percepci místa artikulace nazál, což se pomocí provedeného percepčního testu podařilo – jak je patrné ze srovnání grafů 8a a 8b, hodnocení téže nazály se v některých případech výrazně změnilo pouhou výměnou předcházející slabiky.

Že bezprostřední okolí je v percepci segmentů klíčové, ukázala již řada studií uvedených v Úvodu (kap. 1); tato práce se však od dřívějších studií využívajících podobnou metodu v jednom podstatném ohledu liší. Ve středu pozornosti dříve bývala relativní důležitost různých akustických korelátů místa artikulace – pracovalo se tedy buď se stimuly různým způsobem zbavenými některých akustických vodítek (např. Kurowski a Blumstein, 1984; Malécot, 1956; Repp, 1986), nebo s lepenými stimuly obsahujícími vodítka jdoucí proti sobě (Malécot, 1956), aby se ukázalo, které z nich percepčně převažuje. V této práci však lepené stimuly neobsahovaly záměrně protichůdné akustické informace – naopak, byly slepeny ze slov, která obsahovala vždy [n], a tudíž i okolní segmenty by měly být vlivem koartikulace na [n] „nastaveny“. I přesto bylo změnou kontextu u některých položek dosaženo statisticky významné nejednoznačnosti vnímaného místa artikulace, v případě pseudoslova „rýne“ dokonce k prokazatelnému příklonu k [m] (že zdrojové stimuly takovouto nejednoznačností netrpěly, dokládá graf 8a).

S tradiční teorií lokusu (viz 2.2) nejsou výsledky tohoto experimentu v souladu. Pokud by tranzienty F2 vokálů sousedících s nazálou měly směřovat do jednoho bodu daného místem její artikulace a tím umožňovat její identifikaci, neměla by záměna sousedního segmentu za segment s tranzientem F2 směřujícím do téhož bodu percepci místa artikulace nijak výrazně ovlivnit; to se však děje. Z téhož důvodu zde neuspěje ani žádná jiná teorie tvrdící, že v segmentech sousedících s nazálou je obsaženo jednoznačné akustické vodítko pro určení místa artikulace, nezávislé na kontextu (např. tvar tranzientů či *locus equations*); přinejmenším v některých případech je totiž zjevně možné posluchače „zmást“, ačkoliv příslušná vodítka jsou podle všeho v signálu přítomna. Stejně tak tyto výsledky hovoří proti percepční účinnosti nazálního brumu – ten byl ve zdrojových i lepených stimulech zcela to-

tožný (a neobsahoval ani diskontinuitu charakteristickou pro výstupy plzeňského syntetizéru, který pracuje se segmenty dělenými v polovině - viz 1.1), a přesto byly některé nazály vnímány nejednoznačně. A nepomůžeme si zde ani teoriemi pracujícími s percepční integrací tranzientů a nazálního brumu jakožto dvou nezávislých vodiček (Repp, 1987b), neboť ačkoliv v našem případě tato vodítka zůstávají stejná, percepční dojem se mění.

Nadějně na první pohled vypadá dynamické vodítko toho druhu, jaký nabízí Kurowski a Blumstein (1987), to jest rozdíl mezi spektrem signálu těsně před hranicí nazála-vokál a těsně za ní. Při záměně sousedního segmentu se změní i tento spektrální vztah, což může vést k percepčně odlišnému výsledku. Je tu však několik obtíží. Zaprvé: Kurowski a Blumstein pracovaly s CV slabikami, ale v případě této studie zůstává sekvence CV beze změny. Mění se sekvence VC – a jak poznamenává např. Repp (1987b), přechod mezi vokálem a nazálou je v těchto slabikách méně výrazný než u sekvencí CV (jasně identifikovatelný zlom často chybí), takže vodítko založené na srovnávání spekter kolem tohoto přechodu nepůsobí příliš pravděpodobně. Zadruhé: Ve studii Kurowské a Blumsteinové se toto dynamické vodítko odvědělo pro slabiky obsahující vokály [a, o, u], kdežto v případě vokálů [e] a zejména [i] selhávalo. A jsou to právě sekvence s vysokými předními vokály, které byly v našem experimentu hodnoceny nejednoznačně.¹⁹ Zatřetí: Pokud by se mělo jednat o vodítko invariantní alespoň pro slabiky s určitou užší skupinou vokálů (např. vysokými předními), zůstalo by záhadou, že některá pseudoslova se sekvencí [me] či [i:ne] byla hodnocena jednoznačně, a jiná nikoliv.

Nadějně na první pohled vypadá dynamické vodítko toho druhu, jaký nabízí Kurowski a Blumstein (1987), to jest rozdíl mezi spektrem signálu těsně před hranicí nazála-vokál a těsně za ní. Při záměně sousedního segmentu se změní i tento spektrální vztah, což může vést k percepčně odlišnému výsledku. Je tu však několik obtíží. Zaprvé: Kurowski a Blumstein pracovaly s CV slabikami, ale v případě této studie zůstává sekvence CV beze změny. Mění se sekvence VC – a jak poznamenává např.

¹⁹Že sekvence obsahující přední spíše vysoké vokály se chovají specificky (viz 2.3.2), se tato práce nesnaží potvrdit ani vyvrátit; její rozsah neumožňoval obsáhnout všechny kombinace vokalických kontextů a učinit srovnání. Jak bylo uvedeno na příslušném místě (3.2), do testu byly vybrány takové položky, které podle soudu autorky mohou být percepčně nejednoznačné. Že se v drtivé většině případů jednalo právě o sekvence obsahující [i, i:] (příčemž původním kontextem slabiky [ne] byl vokál [a] či [o]), tvrzení o specifičnosti těchto kontextů sice podporuje, avšak v samotném experimentu mělo úlohu předpokladu, nikoliv testované hypotézy.

Repp (1987b), přechod mezi vokálem a nazálou je v těchto slabikách méně výrazný než u sekvencí CV (jasně identifikovatelný zlom často chybí), takže vodítko založené na srovnávání spekter kolem tohoto přechodu nepůsobí příliš pravděpodobně. Zadruhé: Ve studii Kurowské a Blumsteinové se toto dynamické vodítko odvědělo pro slabiky obsahující vokály [a, o, u], kdežto v případě vokálů [e] a zejména [i] selhávalo. A jsou to právě sekvence s vysokými předními vokály, které byly v našem experimentu hodnoceny nejednoznačně.²⁰ Zatřetí: Pokud by se mělo jednat o vodítko invariantní alespoň pro slabiky s určitou užší skupinou vokálů (např. vysokými předními), zůstalo by záhadou, že některá pseudoslova se sekvencí [me] či [i:ne] byla hodnocena jednoznačně, a jiná nikoliv.

Tímto ovšem nemá být řečeno, že výše uvedená kontextuálně nezávislá akustická vodítka (tranzienty, nazální brum či spektrum na hraně hlásek) nejsou při identifikaci nazálních konsonantů vůbec využívána. Klíčové položky v provedeném percepčním testu byly lepené či manipulované, a jejich akustický průběh tudíž mohl mít takový charakter, s jakým se nelze v přirozené řeči setkat. Není tudíž vyloučeno, že se při konfrontaci s těmito stimuly percepční aparát uchyluje k mechanismům, které v běžné komunikaci využívá spíše výjimečně (například za špatných akustických podmínek). Je však teoreticky i prakticky (zejména s ohledem na konkatenanční syntézu řeči) užitečné se i těmito „defektním“ případům věnovat, neboť ač jsou snad objevené mechanismy aktivovány jen zřídka, přece jimi disponujeme.

Nejvíce jsou přítomné výsledky v souladu se zjištěními Öhmana (1966), který dochází k nutnosti analyzovat sekvence VCV jako celek, neboť spektrální charakteristiky vokálu sousedícího s nazálou (a to ty, které jsou považované za klíčové pro percepci místa artikulace) nejsou nezávislé na kvalitě vokálu, který s nazálou sousedí z druhé strany. Přesně to se zdá potvrzovat provedený experiment – vyřizneme-li slabiku [ne] ze sekvence [ane] a přilepíme ji ke slabice [ri:], vytváříme tím kontext, na který není tato slabika „naladěna“ a výsledkem pak může být percepční dojem, jenž je buď nejednoznačný, nebo dokonce protichůdný našemu očekávání.

²⁰Že sekvence obsahující spíše vysoké přední vokály se chovají specificky (viz 2.3.2), se tato práce nesnaží potvrdit ani vyvrátit; rozsah této práce neumožňoval obsáhnout všechny kombinace vokalických kontextů a učinit srovnání. Jak bylo uvedeno na příslušném místě (3.2), do testu byly vybrány takové položky, které podle soudu autorky mohou být percepčně nejednoznačné. Že se v drtivé většině případů jednalo právě o sekvence obsahující [ɪ, i:] (přičemž původním kontextem slabiky [ne] byl vokál [a] či [o]), tvrzení o specifičnosti těchto kontextů sice podporuje, avšak v samotném experimentu mělo úlohu předpokladu, nikoliv testované hypotézy.

Z výsledků hodnocení manipulovaných položek toho příliš vytěžit nelze. Nárůst percepční nejednoznačnosti oproti zdrojovým stimulům je sice ze srovnání grafů 8a a 8c patrný (například posun v případě pseudoslova „brane“, které před manipulací vykazalo stoprocentní shodu respondentů, je působivý), avšak statisticky významné percepční nejednoznačnosti bylo manipulací dosaženo pouze u slova „bene“²¹ a k jednoznačnému příklonu k [m] se žádné z manipulovaných pseudoslov ani nepřiblížilo. Nelze vyloučit, že další pokusy s manipulacemi by dosáhly jiných výsledků; avšak stejně tak je možné, že zkoumaný jev je omezen na stimuly lepené z různých zdrojů. To by s sebou neslo implikace pro možné využití výsledků této práce, jemuž se věnuje oddíl 5.3.

5.2 Návrh navazujícího výzkumu

Provedeným experimentem bylo pouze potvrzeno, že uvedený jev nastává, dostatečně prozkoumán jím však nebyl. Bylo by především třeba obsáhnout více vokalic-
kých kontextů a zjistit, zdali jde o fenomén specifický pro sekvence obsahující přední a spíše vysoké vokály, či nikoliv. Touž cestou by se také měla zodpovědět otázka, zdali může být percepce místa artikulace takto narušena i v případě, že původní a nový kontext nazály nejsou ve vokalickém prostoru tak daleko od sebe. A dále ovšem čekají na prozkoumání i jiné než vokalické kontexty a s nimi otázka, zdali je tato percepční zvláštnost podmíněna tím, že oba sousední segmenty mají tónový charakter, či nikoliv.

K užitečným výsledkům by pak mohla vést akustická analýza využívaného materiálu. Pokud by se podařilo nasbírat výrazně větší počet promluv vykazujících toto chování, dalo by se pomocí kvantitativních metod pátrat po akustickém parametru, jenž by s tímto chováním koreloval a jenž by jednak mohl přispět k teoretickému pochopení percepce místa artikulace nazál, jednak by byl využitelný prakticky (k tomu více v následující sekci). Možným odrazovým můstkem při tomto pátrání by mohl být koncept *locus equations* (přičemž by musel být brán v úvahu vztah hodnoty tohoto parametru v obou vokálech obklopujících nazálu) či dynamický parametr navrhaný Kurowskou a Blumsteinovou (1984), s výhradami uvedenými výše (5.1).

²¹Vzhledem k předchozím odstavcům stojí za zmínku, že toto je jediné pseudoslovo z oněch pěti, v němž nazále předchází přední vokál.

5.3 Možný přínos uvedených výsledků

Na teoretické rovině se tato práce řadí k těm, které potvrzují neadekvátnost segmentálního přístupu k percepci řeči a ohlašují beznadějnost snahy pátrat po invariantních akustických korelátech percepčních kategorií na segmentální úrovni.

Praktický dopad by výše uvedená zjištění mohla mít v oblasti syntézy řeči, od které ostatně tato práce původně vyšla. Je nepravděpodobné, že by zde zkoumaný jev narušoval percepci přirozené řeči v běžné komunikaci – s případnými nejednoznačnými akustickými vodítky si obvykle posluchač poradí díky sémantice a kontextu (v širokém smyslu toho slova). Navíc se dost možná, jak bylo zmíněno, jedná o fenomén specifický pro řeč syntetickou; vzhledem ke vzrůstajícímu významu řečových technologií jej však tento fakt jistě nestaví mimo oblast výzkumného zájmu.

Výsledky této práce jsou příliš obecné, než aby se nabízely k přímočarému využití v algoritmu konkatenační syntézy. Upozňují však na nutnost brát v úvahu širší kontext difonů, které jsou do výsledné promluvy dosazovány. Do nějaké míry je ohled na kontext samozřejmě v plzeňském algoritmu obsažen už nyní; tato práce pouze odhalila jednu konkrétní oblast, kde současné nastavení může generovat vadné promluvy (i navzdory správně nasegmentované a popsané databázi a bezchybně provedené transkripci). Výhodou korpusu využívaného na KKY v Plzni je, že v databázi jsou uchovávány trifony (viz 1.1) – nemuselo by tedy být příliš obtížné upravit algoritmus syntézy tak, aby upřednostňoval takové difony obsahující nazální konsonant, jejichž původní okolí je podobné kontextu, do něž bude difon zasazen (pro syntézu slova „mlýneček“ by tedy byl spíše vybrán difon [n_e] ze zdrojového slova „klarinetu“ nežli ze slova „Janečka“). Parametrů obsažených v algoritmu je však mnoho a jejich interakce jsou složité – je možné, že by tato restrikce vedla ke zhoršení výsledků syntézy vlivem jiného jevu. Bylo by proto lepší nalézt akustický (nikoliv fonémický) parametr korelující s jevem popsaným v této práci (jak je navrženo v sekci 5.2), a ten následně do algoritmu zahrnout – lze očekávat, že tímto způsobem by byl výběr vhodných difonů omezen méně.

6 Závěr

Tato práce se zabývala vlivem slabiky předcházející nazále na percepci jejího místa artikulace. Navázala tak na rozsáhlý a dlouholetý výzkum věnující se percepci místa artikulace okluziv a přispěla k němu experimentem inspirovaným aktuálně užívanou konkatenací syntézou řeči (a s její pomocí). Provedený percepční test zkoumal, zdali se významně změní percepce nazálního konsonantu, nahradí-li se jeho původní okolí okolím jiným, které však původně sousedilo s toutéž nazálou (fonémicky vzato). Výsledky ukázaly, že k takové změně dojít může, ač to ovšem není pravidlem; a to buď do té míry, že je identifikace místa artikulace pouze znejednoznačněna, nebo i do té, že posluchači jednoznačně volí nazálu, která ve zdrojových slovech nebyla obsažena (to však nastalo v případě pouze jedné položky). Tímto nálezem je opět oslabena hypotéza existence invariantního akustického vodítka pro místo artikulace na segmentální úrovni a potvrzena silná kontextuální závislost vnímání segmentů.

Vzhledem k charakteru stimulů je potenciál využití výsledků této práce největší v oblasti konkatenací syntézy řeči, kde se vlastnosti signálu mění stejně abruptně jako v provedeném experimentu. Vhodný navazující výzkum by mohl směřovat k vylepšení algoritmu syntézy do podoby, která by již neprodukovala problematické stimuly, jakým byl „mlýmeček“ inspirující tuto práci.

Reference

- Audacity Team. (2015). *Audacity(R) [počítačový program]. Verze 2.0.5*. <http://audacity.sourceforge.net>.
- Boersma, P. a Weenink, D. (2015). *Praat: doing phonetics by computer [počítačový program]. Verze 6.0.04*. <http://www.fon.hum.uva.nl/praat/>.
- Clark, J. a Yallop, C. (1995). *An introduction to phonetics and phonology* (2. ed.). Oxford: Blackwell Publishing.
- Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M. a Gerstman, L. (1952). Some experiments on the perception of synthetic speech sounds. *The Journal of the Acoustical Society of America*, 24(6), 579–606.
- Delattre, P. C., Liberman, A. M. a Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *The Journal of Acoustical Society of America*, 27(4), 769–773.
- Diehl, R. L., Lotto, A. J. a Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55, 149–179.
- Fant, G. (1960). *Acoustic theory of speech production*. Haag: Mouton.
- Fant, G. (1969). Stops in cv-syllables. *STL-QPSR*, 10(4), 1–25.
- Farnetani, E. a Recasens, D. (2010). Coarticulation and connected speech processes. In W. J. Hardcastle, J. Laver a F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (2. ed., pp. 316–352). Oxford: Blackwell Publishing.
- Fowler, C. A. a Smith, M. R. (1986). Speech perception as „vector analysis“: an approach to the problems of invariance and segmentation. In J. S. Perkell a D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 123–139). Hillsdale: Lawrence Erlbaum Associates.
- Holt, L. L. a Kluender, K. R. (2002). General auditory processes contribute to perceptual accommodation of coarticulation. *Phonetica*, 57, 170–180.
- Johnson, K. (2005). Speaker normalization in speech perception. In D. B. P. ans R. E. Remez (Ed.), *The handbook of speech perception* (pp. 363–389). Oxford: Blackwell Publishing.
- Johnson, K. (2012). *Acoustic and auditory phonetics* (3. ed.). Oxford: Wiley-Blackwell.
- Kurowski, K. a Blumstein, S. E. (1984). Perceptual integration of the murmur and formant transitions for place articulation in nasal consonants. *The Journal of the Acoustical Society of America*, 76(2), 383–390.
- Kurowski, K. a Blumstein, S. E. (1987). Acoustic properties for place of articulation in nasal consonants. *The Journal of the Acoustical Society of America*, 81(6), 1917–1927.
- Ladefoged, P. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America*, 29(1), 98–104.
- Lehiste, I. a Peterson, G. E. (1961). Transition, glides, and diphthongs. *The Journal of the Acoustical Society of America*, 33(3), 268–277.

- Liberman, A. M. (1957). Some results of research on speech perception. *The Journal of the Acoustical Society of America*, 29(1), 117–123.
- Liberman, A. M., Delattre, P. C., Cooper, F. S. a Gerstman, L. J. (1954). The role of consonant-vowel transition in the perception of the stop and nasal consonants. *Psychological Monographs: General and Applied*, 68(8, Whole No. 379), 1–13.
- Liberman, A. M., Harris, K. S., Hoffman, H. S. a Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5), 358–368.
- Machač, P. a Skarnitzl, R. (2009). *Fonetická segmentace hlásek*. Praha: Epoque.
- Malécot, A. (1956). Acoustic cues for nasal consonants: An experimental study involving a tape-splicing technique. *Language*, 32(2), 274–284.
- Matoušek, J., Tihelka, D. a Romportl, J. (2006). Current state of czech text-to-speech system artic. In P. Sojka, I. Kopeček a K. Pala (Eds.), *Text, speech and dialogue: 9th international conference tsd 2006, lecture notes in artificial intelligence, vol. 4188* (pp. 439–446). Berlin/Heidelberg: Springer.
- Mitterer, H. a Cutler, A. (2006). Speech perception. In E. K. Brown, R. E. Asher a J. M. Y. Simpson (Eds.), *Speech interface design* (pp. 770–782). Oxford: Elsevier.
- Nakata, K. (1959). Synthesis and perception of nasal consonants. *The Journal of the Acoustical Society of America*, 31(6), 661–666.
- Öhman, S. E. G. (1966). Coarticulation in vcv utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America*, 39(1), 151–168.
- Pisoni, D. B. a Remez, R. E. (2005). Introduction. In D. B. Pisoni a R. E. Remez (Eds.), *The handbook of speech perception* (pp. 1–3). Oxford: Blackwell Publishing.
- R Development Core Team. (2016). *R: A language and environment for statistical computing [počítačový program]*. Verze 3.3.2. <http://www.R-project.org>.
- Raphael, L. J. (2005). Acoustic cues to the perception of segmental phonemes. In D. B. Pisoni a R. E. Remez (Eds.), *The handbook of speech perception* (pp. 182–206). Oxford: Blackwell Publishing.
- Recasens, D. (1983). Place cues for nasal consonants with special reference to catalan. *The Journal of the Acoustical Society of America*, 73(4), 1346–1353.
- Repp, B. H. (1986). Perception of the [m]-[n] distinction on cv syllables. *The Journal of the Acoustical Society of America*, 79(6), 59–85.
- Repp, B. H. (1987a). On the possible role of auditory short-term adaptation in perception of the prevocalic [m]-[n] contrast. *The Journal of the Acoustical Society of America*, 82(5), 1525–1538.
- Repp, B. H. (1987b). Perception of the [m]-[n] distinction on vc syllables. *The Journal of the Acoustical Society of America*, 83(1), 237–247.
- Samuel, A. G. a Pitt, M. A. (2003). Lexical activation (and other factor) can mediate compensation for coarticulation. *Journal of Memory and Language*, 48, 416–434.

- Skarnitzl, R., Šturm, P. a Volín, J. (2016). *Zvuková báze řečové komunikace*. Praha: Karolinum.
- Summerfield, Q., Haggard, M., Foster, J. a Gray, S. (1984). Perceiving vowels from uniform spectra: Phonetic exploration of an auditory aftereffect. *Perception & Psychophysics*, 35(3), 203–213.
- Sussman, H. M., McCaffrey, H. A. a Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *The Journal of the Acoustical Society of America*, 90(3), 1309–1325.
- Sussman, H. M. a Shore, J. (1996). Locus equations as phonetic descriptors of consonantal place of articulation. *Perception & Psychophysics*, 58(6), 936–946.

7 Přílohy

Příloha 1

Obsahem této přílohy je zvukový materiál připojený k elektronické verzi práce – syntetizované slovo „mlýneček“ a dvě slova zdrojová: „Jelínka“ (zdroj difonu [i:_n]) a „Janečka“ (zdroj difonu [n_e]). Jedná se o tentýž hlas, který byl použit v percepčním testu.

Příloha 2

Obsahem této přílohy je zvukový materiál připojený k elektronické verzi práce – syntetizované slovo „mlýneček“ a dvě slova zdrojová: „líný“ (zdroj difonu [i:_n]) a „donekonečna“ (zdroj difonu [n_e]). Jedná se o další mluvčí z korpusu plzeňského syntetizéru (Iva, kód /artic_images/spkr_KI.rev611.img).

Příloha 3

Výřez z tabulky (zahrnující 18 z celkového počtu 2240 položek) ilustruje formu zpracovávaných dat. Sloupec „subject“ obsahuje kód posluchače, „stimulus“ název položky, „response“ zvolenou odpověď (kde „left“ značí variantu s [n] a „right“ variantu s [m]) a „manipulation“ typ stimulu (zdrojový, lepený, manipulovaný). Následují klíčové úseky skriptu použitého k analýze výsledků.

subject	stimulus	response	manipulation
tom_pro	čí-ná	left	lepene
tom_pro	čí-ná	right	lepene
tom_smr	čí-ná	left	lepene
tom_smr	čí-ná	left	lepene
ver_hom	čí-ná	right	lepene
ver_hom	čí-ná	right	lepene
ver_man	čí-ná	right	lepene
ver_man	čí-ná	right	lepene
vla_pta	čí-ná	right	lepene
vla_pta	čí-ná	right	lepene
ane_mix	ameč	left	manip
ane_mix	ameč	left	manip
ann_kuc	ameč	left	manip
ann_kuc	ameč	right	manip
eva_ull	ameč	left	manip
eva_ull	ameč	left	manip
fil_rut	ameč	left	manip
fil_rut	ameč	left	manip

Výpočet průměrných hodnocení a vytváření grafu pro skupinu lepených stimulů:

```
alpha <- 0.05 / 15 # Bonferroniho korekce
mean_tb_boot <- function(x) {
  set.seed(1)
  return(mean_cl_boot(x, conf.int = 1-alpha))
}
tab_lepene <- subset(tab, subset = (tab$manipulation == "lepene"))
tab8 <- tab_lepene %>%
  mutate(response = if_else(response == "left", -1, 1)) %>%
  group_by(subject, stimulus) %>%
  summarise(response = mean(response), RT = mean(reactionTime))
ggplot(tab8, aes(x = reorder(stimulus, response), y = response)) +
  geom_abline(intercept = 0, slope = 0, color = "darkgray") +
  stat_summary(fun.data = "mean_tb_boot", fatten = 2) +
  stat_summary(fun.data = "mean_tb_boot",
              geom = c("errorbar", "pointrange"), width = .2) +
  expand_limits(y=c(-1,1)) +
  labs(x="stimul", y="průměr odpovědí", title="b) Lepené stimuly") +
  theme(axis.text.x = element_text(angle = 60, vjust = 0.6, hjust=0.4))
```

Vytvoření dílčích tabulek pro srovnání lepené položky "rý-ne2" se zdrojovou položkou "pone2" a provedení párového Wilcoxonova testu:

```
tabX <- tab %>%
  mutate(response = if_else(response == "left", -1, 1)) %>%
  group_by(subject, stimulus) %>% summarise(response = mean(response))
tabX <- ungroup(tabX)
pone2 <- subset(tabX, (tabX$stimulus == "pone2"))
ryne2 <- subset(tabX, (tabX$stimulus == "rý-ne2"))

wilcox.test(x = pone2$response, y = ryne2$response,
            alternative = "two.sided", paired = TRUE,
            exact = FALSE, correct = TRUE, conf.level = 0.95)
```