Univerzita Karlova v Praze
Filozofická fakulta
Fonetický ústav

Diplomová práce

Kristian Urban

Syntéza znělostních charakteristik českých exploziv a jejich předběžné percepční ověření

Synthesis of the voicing characteristics of Czech plosives and their preliminary perceptual verification.

Praha 2013

Vedoucí práce: doc. Radek Skarnitzl, Ph.D.

Poděkování: Rád bych poděkoval svému vedoucímu práce, doc. Radku Skarnitzlovi, Ph.D., za jeho pomoc, ochotu, trpělivost a přátelský přístup. Bez jeho pomoci by tato práce nikdy nevznikla. Děkuji.

Prohlašuji, že jsem diplomovou práci vypracoval samostatně, že jsem řádně citoval 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 27. srpna 2013 *podpis*

Abstrakt: Tato práce se zabývá znělostními charakteristikami českých exploziv. V teoretické části je nastíněna funkce hlasivek a jejich podíl na znělosti hlásek. Dále je přiblížena historie, dělení a aplikace řečových syntetizátorů. Právě syntetizéry umožňují kontrolovanou manipulaci jednotlivých parametrů řečového signálu. V praktické části jsou pomocí syntetizéru HLsyn syntetizovány jednotlivé české explozivy s různými znělostními charakteristikami. Na základě předběžného percepčního ověření je pak vyhodnoceno, jak je posluchači znělost vnímána.
Klíčová slova: znělost, explozivy, syntéza, HLsyn, percepční test

Abstract: This paper will discuss voicing characteristics of Czech plosives. In the theoretical part of the paper, vocal cords are described as well as their participation on voicing. Next, a brief history, division, and application of voice synthesisers is discussed. Synthesizers allow the user to manipulate individual characteristics of any speech signal. In the practical part of the paper, HLsyn is used to synthesize individual Czech plosives with various voicing characteristics. Perceived voicing is then evaluated based on preliminary perceptual verification.
Keywords: voicing, plosives, synthesis, HLsyn, perception test

Obsah

1. Úvod

Byl jsem ještě malý chlapec, když jsem poprvé žasnul nad tím, jak může počítač komunikovat s lidmi pomocí hlasu. Jistě, tenkrát jsem byl neznalý celé problematiky a vůbec obtížnosti provedení, nicméně jsem tím byl fascinován a bez pochyb to byla jedna z věcí, která mě tenkrát přitáhla k seriálu Star Trek. Samozřejmě, že se jednalo o fikci, a v té míře, nad kterou jsem tehdy žasl, je to fikce i dnes. Rozdíl je ten, že dnes to již nevidím jako něco magického a tajemného, ale vidím konkrétní problémy, možná řešení a na obecné úrovni i možnosti celé komunikace mezi člověkem a počítačem.

Stroje fascinovaly člověka snad již od doby, kdy první komplexní stroj sestavil. V tu chvíli začal snít o (s nadsázkou) neomezených možnostech, které člověku mohou stroje poskytnout. Přeskočíme několik staletí vynálezů a zastavíme se pro tuto chvíli v době poměrně nedávné: na počátku dvacátého století. Konkrétně v roce 1938, kdy lze datovat dokončení prvního programovatelného počítače. Ten sestavil německý inženýr Konrad Zuse (Rojas & Hashagen, 2002, str. 237). Z dnešního pohledu by možná mnozí s nadsázkou pochybovali o tom, jestli tehdejší sálové počítače vůbec počítači byly. Byly to z dnešního pohledu obrovské zprvu mechanické stroje, se kterými se obtížně komunikovalo. Obecně by se dalo shrnout, že snahy v oblasti komunikace člověk – počítač, nebo chceme-li více obecně člověk – stroj, jdou buď po ose „já se přizpůsobím počítači“ nebo „počítač se přizpůsobí mně“. Pro odborníky není problém komunikovat se stroji v jejich vlastním, programovacím, jazyce pomocí příkazové řádky. Běžný uživatel se jen těžko může touto měrou přizpůsobit počítači a tak bylo počítači vtisknuto do vínku tzv. grafické prostředí (zkr. GUI). To je ve zkratce jakási nadstavba nad jazyk počítače, která je uzpůsobena pro snadnou visuální komunikaci s uživatelem.

GUI je však stále jen určité uměle vytvořené médium, takový mezistupeň, skládající se z ikon a symbolů. Velký krok pro přiblížení člověka a počítače přišel v roce 1960, kdy Douglas Engelbart vynalezl vstupní zařízení, které umožňovalo pohybovat kurzorem a zvolit

požadovanou funkci přímo na obrazovce (O'Regan, 2012, str. 187). Toto zařízení se vžilo pod názvem myš. Dnes se podobná revoluce děje na frontě mobilních zařízení a dotykových obrazovek, kdy myš začíná být nahrazována mnohem přirozenějším přímým dotykem na obrazovku.

Přesuneme-li se však opět v úvahách o krok zpět, musíme dojít k závěru, že s počítačem stále komunikujeme pomocí grafiky a písma. Jakkoliv snadné a dokonalé nám dnes může toto ovládání připadat, z hlediska vývoje techniky a především z hlediska toho, jak člověk přirozeně komunikuje, se jedná stále o jistou „berličku“ v užívání počítačů. Nejpřirozenější a zároveň nejsnazší způsob komunikace pro člověka stále zůstává řeč (Holmes & Holmes, 2001, str. 1).

1.1. Vztah jazyka, písma, myšlenek a strojů

Pro větší pochopení problematiky komunikace člověka se strojem je zapotřebí se zamyslet nad komunikací jako takovou. Pro většinu lidí může ideální „předpisovou“ podobu jazyka představovat písmo. Tedy něco kodifikovaného, srozumitelného, exaktního a dokonalejšího než mluvená řeč. Faktem ovšem zůstává, že mluvené podobě řeči se naučíme dříve než podobě psané. Rovněž se v mluvené podobě snáze zachytí jemnější detaily v komunikaci než strohým psaným projevem. Chceme-li skutečně vyřešit nějakou neshodu (či obecně vzato jakýkoliv komunikačně náročnější cíl), je pohodlnější a efektivnější zvolit cestu mluvené formy řeči, ideálně přímo z očí do očí. Buď jak buď, psaný či mluvený projev, v tom či v onom jazyce, cíl je pouze jeden – sdělit myšlenku, komunikovat. Je dobré si na tomto místě uvědomit, že slova, věty a celé projevy jsou vlastně jen prázdný nosič, za kterým se teprve informační hodnota schovává. Slova jako taková jsou pouze prostředek, nikoliv cíl. Mnohdy stejnými slovy sdělujeme informaci různou, či naopak různými slovy informaci stejnou. Čtenáři již v tuto chvíli musí být jasné, že reálná komunikace člověka a stroje skrývá problémy nejen v rovině formy, ale i v rovině obsahu. Můžeme stroj naučit reprodukovat či rozpoznat mluvené slovo, ale naučit ho chápat význam slova, to přináší mnoho dalších

problémů. Jejich řešením se zabývají tzv. dialogové systémy (Kopeček, 2013). Jedním z pracovišť, které tuto problematiku zkoumá, je Laboratoř vyhledávání a dialogu Fakulty informatiky Masarykovy univerzity v Brně.

Zde stojí za zmínku jeden z nedávných směrů výzkumu. Společnosti Microsoft a Facebook se do svých produktů snaží zařadit funkci tzv. přirozeného jazyka. V praxi by to znamenalo, že například v tabulkovém procesoru Excel bude možno namísto zažitých vzorců používat zadání typu „sečti všechny náklady spojené s dopravou za měsíc březen“ a tabulkový procesor si sám vyhledá potřebná data v řádcích a sloupcích (Urban, 2013). Populární sociální síť Facebook chystá vyhledávání na podobném principu, kdy bude možné zadat do vyhledávání dotaz typu „najdi mi všechny uživatele, kteří bydlí v Praze, mluví anglicky a zajímají se o filmy“ (Václavík, 2013).

Převážná většina lidské komunikace se odehrává pomocí mluveného slova. Samozřejmě existuje nonverbální informační kanál, který nese neuvěřitelné množství informací a je velmi důležitý, ale to, kam pojedete s rodinou na dovolenou nebo v kolik hodin odjíždí vlak, se pomocí gest nedozvíte (pochopitelně s odhlédnutím od znakové řeči). I přes neuvěřitelné množství textu, které člověk za běžný den přečte a vstřebá, je pro nás verbální komunikace to nejdůležitější (Matoušek, Studijní materiál ke kurzu, 2010). Termínem verbální se zde míní skutečně mluvenou řeč.

Jakkoliv se nedá budoucí vývoj odhadnout, tak lze předvídat, že poptávka po mluvené interakci se stroji bude čím dál větší, až se nakonec stane primárním komunikačním kanálem mezi člověkem a počítačem pro běžné záležitosti každodenního života. K plynulé mluvené komunikaci však vědcům zbývá ještě několik důležitých a velkých kroků a byla-li by možnost použít pohádkové sedmimílové boty, jistě by nebylo od věci si je obout a vykročit.

Komunikace s počítačem má několik oblastí, které nemusí být běžnému člověku na první pohled zřejmé. Čistě pro přehled je lze ve stručnosti uvést. Za prvé se jedná o rozpoznání lidské řeči. Kolegové z Fakulty informačních technologií VUT v Brně by mi jistě dali za

pravdu, kdybych řekl, že universální rozpoznávač souvislé lidské přirozené řečí je něco, co se velmi těžko sestavuje. Jak se říká: každý jsme jiný. Přičemž počítače jsou stále jen stroje, které nemají možnost lidské abstrakce. Pokud si například představíte rozdíl mezi výslovností obyvatel Prahy a obyvatel Ostravy, pak vězte, že pro spolehlivé rozpoznání mluveného slova počítačem mohou být problém i dva obyvatelé Prahy vůči sobě.

Dalším krokem ve výše popisované komunikaci mezi člověkem a počítačem by ideálně měla být intelektualizace řečeného a vygenerování obsahu výpovědi (v konkrétním případě pak většinou odpovědi na danou otázku). Tento krok jde již za hranu fonetických oborů a sám o sobě je to velmi komplexní lingvistický problém s nemalým přesahem do matematiky, filosofie a zřejmě i sociálního inženýrství. Za první větší rozšíření podobných technologií mezi veřejnost lze nejspíše považovat pomocníka Siri společnosti Apple (Apple Inc., 2013). Ten je schopen s uživatelem komunikovat na dnešní poměry na poměrně vysoké úrovni. Jeden příklad za všechny je otázka „Co si mám vzít na sebe?“, která vyvolá odpověď z oblasti předpovědi počasí. Kouzelný na tom je právě fakt, že Siri dle všech předpokladů nemá předem generované typy hesel, na které ploše reaguje, ale nad zadáním „přemýšlí“. Hlubší fungování systému je bohužel kryto jako průmyslové tajemství, a tak zde není možno pod jeho pokličku nahlédnout. Komunikace na o něco nižší úrovni už funguje pár let například v automobilech v podobě zabudovaného hlasového vytáčení (a samozřejmě i přímo ve většině dnešních mobilních telefonů). Dříve si člověk pro hlasové vytáčení musel pořídit svou hlasovou nahrávku povelu, který pro telefon znamenal vzor hlasového pokynu pro volání na dané číslo. Jinými slovy: pokud jste si u položky „Jana Nováková“ v telefonním seznamu nahráli hlasový povel „maminka“, pak telefon reagoval na hlasový povel „maminka“. Dnes se do vyšších modelů aut jednotlivých automobilek, například do vozů Škoda Octavia druhé generace a novějších, které jsou cílené na český trh, dodává autorádio schopné se bezdrátově pomocí technologie Bluetooth spojit s mobilním telefonem uživatele a zjistit si seznam kontaktů. Uživateli pak stačí vyslovit bez puštění volantu „Volej Jan Novák mobil“.

Autorádio následně potvrdí zadání větou „Chcete volat Jan Novák mobil?“. Po potvrzení slovem „Ano“ ze strany uživatele dojde k vytočení požadovaného čísla. Důležitá změna oproti výše uváděnému systému hlasového vytáčení je fakt, že již není zapotřebí předem namlouvat hesla. Záleží čistě na textovém poli „jméno kontaktu“ v telefonním adresáři mobilního telefonu. To znamená, že autorádio dokáže nejen rozpoznat hlasový příkaz ke každé položce v seznamu kontaktů uživatele, ale každou položku umí i vyslovit v rámci potvrzovací fráze, která je plně syntetizovaná. Zároveň je nutné poznamenat, že tento konkrétní systém je skutečně použitelný v životě, minimálně za běžného městského provozu. Lze si však představit, že za vyšších rychlostí už bude použití problematičtější a to nejen díky aerodynamickému hluku z vně vozu. Podobným hlasovým systémem umožňuje vybavit své vozy téměř každá automobilka, v českém jazyce se však jedná spíše o výjimku.

Posledním krokem popisované komunikace je vygenerování zvukového signálu, který bude posluchačem vnímán jako lidský hlas, samozřejmě alespoň v tom ideálním případě. Tento poslední krok je to, co bude předmětem diskuse a zkoumání této práce. Syntéza řeči je tedy jeden z mnoha kroků, který přiblíží komunikaci člověka se strojem více člověku. Ponechme již ostatní zmíněné oblasti stranou a pojďme se nyní věnovat čistě řečové syntéze.

2. Syntetizéry a jejich historie

Zcela jistě čtenáři na mysli vyvstane i otázka, proč by měl umět počítač mluvit. Před každým vynaložením energie určitým směrem je dobré se zeptat, proč to či ono dělat. Již dnes se v běžném životě setkáme s automatizovanými hlasovými systémy například v metru, na nádraží, ale i v automobilové navigaci. Ostatně takový příklad z praxe byl již blíže nastíněn v jednom z předešlých odstavců. Hlasové systémy lze jistě zlepšovat, ale použitelné a kvalitní úrovně již dosáhly a zřejmě již není zapotřebí je dále výrazně vylepšovat, pouze je doladit. Naproti tomu pole jako výuka jazyků, pomoc postiženým lidem nebo zábavní průmysl (především v podobě počítačových her) jde stále výrazně zlepšovat a je to rozhodně na místě. Za zmínku stojí fakt, že strojem generovaná řeč je prozatím považována za něco méně kvalitního řeč lidská. Minimálně se traduje hypotéza, že za ztížených poslechových podmínek klesá srozumitelnost syntetizovaného projevu strměji než u projevu lidského mluvčího (Tatham & Morton, 2005, str. 5). Na druhou stranu se objevují názory, že jsou situace, ve kterých lze naopak využít uměle generované řeči k lepším výsledkům než s použitím živého mluvčího. Jedním z podstatných faktorů v tomto ohledu je tempo produkce řeči. Zatímco člověk má své limitace a za určitou hranicí už z hlediska tempa není schopen srozumitelné produkce, stroje mají oproti tomu teoretický práh daleko výše. Nevýhoda nepřirozené produkce řeči (i v závislosti právě na zmíněném vyšším tempu) se tak může paradoxně stát předností a znamenat svým způsobem i zlepšení nad úroveň živého mluvčího (Tatham & Morton, 2005, str. 2). Příkladem budiž například hlasový informační systém na nádražích nebo strojové předčítání knihy, kdy lze upravit tempo řeči tak, že dochází k úspoře času, nikoliv však ke ztrátě informace, jak by docházelo u lidského mluvčího, který by zvýšeného tempa musel dosahovat na úkor kvality projevu.

Posuňme se v lidské historii opět o několik staletí zpět. Daleko před vynález počítače, tištěného spoje a jiných elektrotechnických vymožeností dneška. Snaha uměle generovat lidskou řeč tu byla již v roce 1779 v podání Christiana Kratzensteina. Známější příklad

následoval v roce 1791 v podání von Kempelena a jeho mluvícího stroje (Matoušek, Studijní materiál ke kurzu, 2010). Jejich úsilí poměrně přirozeně plynulo ze snahy co nejlépe napodobit vokální trakt člověka. Ukázalo se však, že přesná fyzikální imitace lidského hlasového ústrojí k dobrým výsledkům nevede a bude nutné se vydat jinou cestou. Počátkem dvacátého století tedy začaly přicházet první elektronické syntetizéry. Konkrétně to byl H. Dudley v roce 1936 se syntetizérem Voder. S nástupem počítačové techniky pak snahy uměle tvořit lidský hlas přirozeně přešly právě na toto pole a od 60. let dvacátého století známe tzv. digitální syntetizéry (Matoušek, Studijní materiál ke kurzu, 2010). Ty jsou z dnešního pohledu tím nedůležitějším. Jedním ze směrů je tzv. artikulační syntéza, dále se pak pracuje na tzv. formantových syntetizérech, které byly později doplněny konkatenačními syntetizéry, a na závěr si ještě zmínku zaslouží systémy TTS (text-to-speech; převod psaného textu na řeč), které lze vnímat spíše jako nadstavbu nad již existující syntetizéry.

2.1. Od srozumitelnosti k přirozenosti

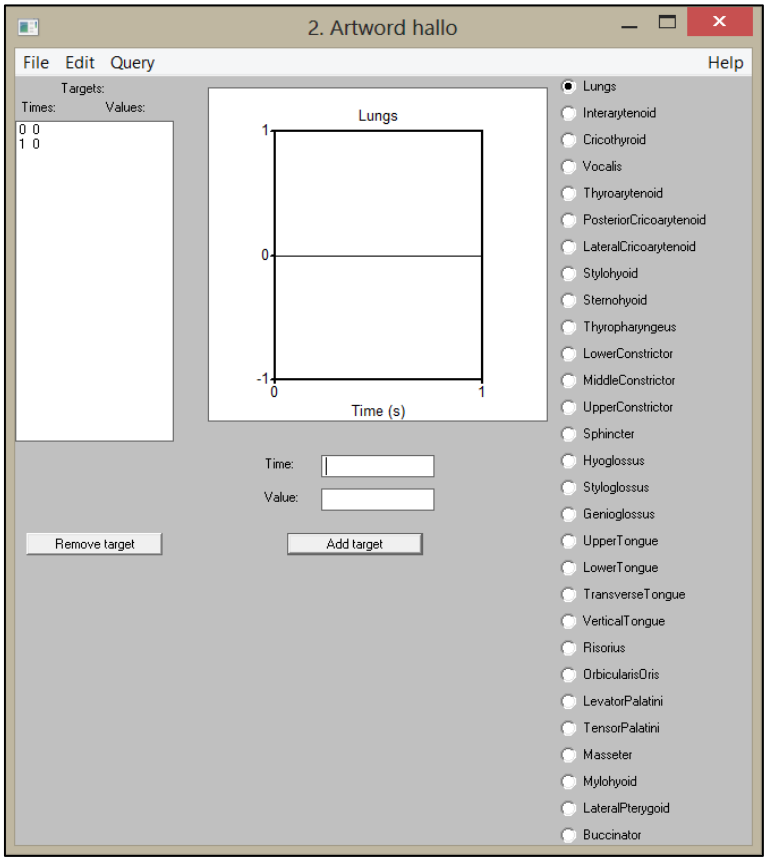
Ještě než budou blíže popsány jednotlivé typy syntetizérů, je nutné se alespoň okrajově zmínit o tom, jaká jsou kritéria hodnocení z pohledu kvality syntézy. Co se hodnotí dnes, není to samé, co se hodnotilo dříve. Kritéria hodnocení se mění spolu s tím, jak daleko je zkoumání v té které oblasti. Nejprve se hodnocení syntetizérů odvíjelo od toho, zda je vůbec syntetizovaným slovům rozumět (Tatham & Morton, 2005, str. 2). K porozumění prvních syntetizovaných slov musel člověk doslova napínat uši. Později, když už byly nalezeny cesty jak syntetizovat řeč srozumitelně, se začal klást větší důraz na přirozenost promluvy. Hodnotit přirozenost se však v praxi ukazuje jako velmi komplikovaný úkol. Těžko se hledá konsensus toho, čím je vlastně ona přirozenost v hlase reprezentována. Nicméně i tak jsou pokusy přirozenost popsat. Jako příklad uveďme hodnocení přirozenosti, které představili Sluijter a kolektiv (Sluijter, a další, 1998, str. 2). Pouze pro představu zde uvedeme některá z jedenácti navrhovaných kritérií: celková kvalita, jednoduchost porozumění, srozumitelnost,

tempo promluvy, příjemnost hlasu, živost a přátelskost. Je patrné, že většina parametrů je subjektivního charakteru, ale to je ve skutečnosti celé hodnocení přirozenosti čehokoliv.

2.2. Artikulační syntéza

Artikulační syntéza pracuje na podobném principu jako původní fyzikální modely lidského hlasového ústrojí. Na rozdíl od nich ale hlasové ústrojí neexistuje jako hmatatelný model, ale pouze jako virtuální simulace v počítači. Artikulační syntéza pracuje s přesným matematicko-fyzikálním modelem, který zohledňuje neurofyziologické a biomechanické procesy spojené s produkcí řeči (Tatham & Morton, 2005, str. 23). Ačkoliv je to ze všech dnes používaných metod ta nejméně rozvinutá, paradoxně může poskytovat nejlepší výsledky co do hlediska přesnosti produkce (Tatham & Morton, 2005, str. 23). Čím dokonalejší virtuální model hlasového ústrojí, tím lepší může být výsledek. V ideálním případě lze dosáhnout 100% věrné simulace lidského mluvčího. Výhoda systému je však úzce spjata i s jeho negativy, která, alespoň prozatím, převažují nad pozitivy. Matematicky a fyzikálně popsat věrnou kopii (nebo jen aproximaci) lidského hlasového ústrojí je velmi obtížné. Pro skutečně detailní popis a syntézu se do výpočtů zahrnuje například i adekvátní model pro motorické funkce muskulatury nejen orální ale i pulmonické, včetně všech souvztažností. Dále by bylo zapotřebí zahrnout model biomechanický, aerodynamický a akustický. V ideálním případě by pak měl model obsahovat matematické formule pro všechny typy fonací.

Pro bližší představu toho, jak náročné je používat artikulační syntetizér, si uveďme několik parametrů, které se musí do syntetizátoru reálně zadávat. Při artikulační syntéze ve fonetickém programu Praat (Boersma & Weenink) pracuje uživatel s celkem dvaceti devíti parametry, mezi kterými jsou parametry pro plicní činnost, činnost svalů hrtanu, detailní popis polohy jazyka, poloha měkkého patra, rtů a další, přičemž u každého parametru se musí nastavit cílové hodnoty pro klíčové momenty ve zvolený čas. Bližší pohled na pracovní okno artikulační syntézy viz obrázek (Obrázek 1). Prakticky se tedy pracuje s hodnotami, které odráží skutečné fyziologické nastavení vokálního traktu.

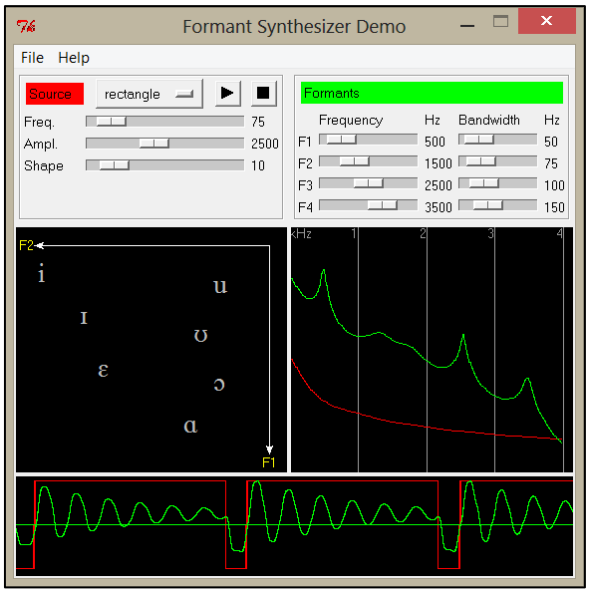


Obrázek 1 Nastavení parametrů artikulační syntézy v programu Praat.

2.3. Formantová syntéza

Formantová syntéza vychází z předpokladu, že k úspěšnému napodobení lidské řeči stačí simulovat formanty. Pro formantovou syntézu tedy v praxi stačí popsat proces vzniku cílových formantů. Toho je docíleno tak, že se použije model filtru a zdroje (Styger & Keller, 1994, str. 111). Zjednodušeně se tak vlastně kopíruje princip produkce řeči lidského vokálního traktu. Hlasivky jsou reprezentovány zdrojem buzení a prostory nad hrtanem pak filtrem. Nadhrtanové dutiny v praxi fungují jako zvukový filtr, který některé frekvence tlumí a jiné nikoli. Vznikají tak zmíněné formanty – spektrální vrcholky v hlasovém spektru. Pro představu viz obrázek (Obrázek 2).

Zásadním rozdílem oproti artikulačnímu syntetizéru je fakt, že parametry, které se zadávají do formantového syntetizéru, mají velmi blízko k reálným hodnotám výsledného zvuku. Nepopisuje se tak tedy detailní nastavení mluvidel, ale výsledné akustické hodnoty jako frekvence formantů, jejich amplituda a další. Nutno podotknout, že práce s formantovým syntetizérem je výrazně méně náročná než s artikulačním. Rovněž výpočetní náročnost je u formantové syntézy nižší. Pro srovnání výpočetní náročnosti lze odkázat na nápovědu programu Praat (Boersma & Weenink). Tam se uvádí, že při artikulační syntéze jednoduchého shluku vokál konsonant vokál o celkové délce 0,5 vteřiny výpočet zabere přibližně šest vteřin. Ovšem ještě za předpokladu, že máme moderní počítač. Na počítači z roku 1997 by syntéza trvala přibližně pět minut. Na tak krátký úsek je však i několik vteřin příliš dlouhá doba pro praktické využití například v mobilních zařízeních. Pokud výpočet artikulační syntézy dříve trval v řádu minut a dnes v řádu vteřin, pak formantovou syntézu lze považovat za okamžitou. Okamžitou odezvu si lze například vyzkoušet v demu formantové syntézy Jonase Beskowa (Beskow, 2013).



Obrázek 2 Ukázka jednoduchého formantového syntetizéru z dílny Jonase Beskowa. V pravém čtvercovém okně je možné vidět formanty (spektrální vrcholky) neutrálního vokálu [ə].

2.4. Konkatenační syntéza

Po artikulační a formantové syntéze přišel poměrně zajímavý krok v podobě tzv. konkatenačního syntetizéru. Vzhledem k širokým možnostem využití a jasným přednostem je v dnešní době tento způsob syntézy nejrozšířenější (Schwarz, 2000, str. 1). Zatímco výše popisované modely pracovaly s parametry a zaměřovaly se buď na fyziologický nebo akustický popis signálu, konkatenační syntetizér se vydal zcela jinou cestou. Pracuje s databází skutečných řečových signálů lidského mluvčího. Zajímavý je na tomto přístupu

fakt, že namísto obtížného definování propracovaných pravidel pro syntézu, jsou pravidla vytahována přímo z reálných dat v databázi. S ohledem na překvapivě vysokou přirozenost syntetizovaného hlasu se lze domnívat, že právě tento druh syntézy je použit ve dříve zmíněném systému hlasového vytáčení v automobilech značky Škoda.

Než bude princip konkatenační syntézy vysvětlen blíže, je ještě nutné se zastavit u termínu „segment“, který se bude v této kapitole vykytovat často. Za segment lze z technického hlediska označit prakticky jakkoliv velký řečový úsek. Intuitivně by člověk jako segment volil nejspíše hlásku (pak se segment označuje jako fon) či slabiku. V praxi se však s velkým úspěchem používá i segment o délce dvou, respektive tří hlásek (difon, respektive trifon). Bylo-li by to žádoucí, lze za segment označit i jednotku menší než jednu hlásku. Z toho vyplívá, že je nutné vždy segment jako pojem přesně definovat.

Konkatenační systémy pracují s řetězením vzorků řečových signálů, které vybírají z předem zpracované databáze. Ta obsahuje všechny možné kombinace segmentů v daném jazyce. Segmenty jsou získávány nahráváním přirozeného projevu jednoho mluvčího. Ten je požádán, aby na mikrofon namluvil zadaný promluvový úsek. Teprve následně je techniky cílený segment označen a vložen do databáze. Segmenty tedy nejsou nahrávány individuálně, avšak v přirozeném okolí v rámci širší promluvy. V současné době se nejvíce využívá difonů, ale byly užívány i trifony a pokusy byly prováděny i se samotnými fony či poloslabikami (Tatham & Morton, 2005, str. 28). Pro příklad můžeme zmínit, že pro francouzštinu se uvádí, že pro plné pokrytí databáze je zapotřebí zhruba 1200 difonů, což představuje zhruba tříminutový souvislý projev (Dutoit, 2001, str. 187). Teoreticky by šlo brát i jednotlivá slova, ale problém by byl následující. Je možné zjistit pro daný jazyk všechny možné hlásky, slabiky nebo i kombinace hlásek v segmentu difon/trifon. Množství takových jednotek bude sice veliké, ale dopočitatelné a konečné. Čím delší ale segment bude, tím se množství možných jednotek v databázi zvyšuje. Je prakticky nemožné určit všechna slova daného jazyka (včetně deklinací, atp.) a zahrnout je do databáze. Tím spíše, že je nutné mít v databázi vzorky

z různých pozic ve větě, což je důležité například pro věty tázací. Teoreticky možné by to bylo pro omezené množství slov, které by pro syntézu bylo vyžadováno, či pro jazyk o malém („konečném“) počtu slov. V praxi lze za konkatenační syntézu na úrovni celých slov považovat hlasový informační systém na nádraží, kde jsou z databáze řetězena celá slova. Zjednodušený příklad toho, jak by mohl takový systém vypadat, je uveden v tabulce níže (**Error! Reference source not found.**).

Jak už bylo řečeno, databáze konkatenačních systémů jsou tvořeny řečovými segmenty.

Osobní vlak	přijede na	první	kolej	první	nástupiště v	jednu	hodinu	a	jednu	-	minut.
Spěšný vlak		druhou		druhé		tři	hodiny		patnáct	-	minuty.
Rychlík		třetí		třetí		pět	hodin		dvacet	pět	minutu.
Vlak inter city		čtvrtou		čtvrté		osm			třicet	šest	

Tabulka 1 Ukázka možného řetězení slov veřejného hlášení na železniční stanici.

Pro tuto chvíli předpokládejme, že se jedná o difony. Pro obecné pochopení fungování konkatenačního syntetizéru však velikost segmentu není důležitá. Jelikož v izolované formě není difon přirozenou součástí jazyka, není možné mluvčího nechat přednést samotné difony a myslet si, že tím je práce hotova. Musí se pořídit nahrávky všech požadovaných difonů ve spojité řeči. To ovšem zpětně vyvolává problém toho, jak určovat hranice segmentů. To je ve spojité řeči obecně problém. Laická představa, že mluvené slovo se skládá z jasně oddělitelných „písmenek“ je bohužel pro fonetiky mylná. Mluvený spontánní projev je ve skutečnosti spojitý zvukový signál, který má s oddělitelnými písmenky na papíře mnohdy společného jen velmi málo. Částečně se obtíže se segmentací řeší tím, že se používají difony či poloslabiky. Vychází se totiž ne z toho, jak je řeč tvořena, ale jak je vnímána. Při vhodném výběru segmentů je difony možné za sebe napojit, aniž by napojení bylo percepčně rušivé. Pokud totiž pochází ze stejného okolí, pak je často možné je napojit za sebe bez dalších úprav.

Právě o vhodný výběr kandidátů se stará algoritmus zvaný „unit selection“ (Schwarz, 2000, str. 1). Ne vždy je ale situace ideální a musí se napojit segmenty, které pochází z různého okolí. Matoušek uvádí, že nejvhodnější pro přirozené přechody mezi segmenty je vybírat nejdelší možný segment z databáze (Matoušek, Automatic Segmentatio of Parasitic Sounds in Speech Corpora for TTS Synthesis, 2010, str. 369). O samotné napojení vybraných segmentů se stará tzv. konkatenační algoritmus. Je zajímavé, že proti sobě v praxi mohou jít dvě protichůdné síly. K větší přirozenosti při napojení segmentů dochází paradoxně v databázi, kde není kladen důraz na hyperartikulaci. Pokud by však vytvořena na podkladě hyperartikulace databáze byla a k dispozici by byl zároveň velmi kvalitní konkatenační algoritmus, docházelo by ke snadnému napojení segmentů, ale s určitým negativním dopadem na přirozenost (Dutoit, 2001, str. 195).

Hranice hlásek v segmentu v databázi se značí jak vnitřní tak vnější. Jinými slovy: jsou značeny hranice počátku a konce segmentu, ale i jednotlivých hlásek, ze kterých je segment tvořen (je-li samozřejmě segment tvořen více než jednou hláskou). Segmenty tedy mají i své vnitřní hranice. Značení hlásek uvnitř segmentu je potřebné pro pozdější automatizovanou úpravu před samotným generováním řeči. Nejjednodušší příklad je změna tempa řeči. Konec promluvového úseku má nižší tempo oproti začátku úseku. Snížení tempa o například 20% však neznamená rovnoměrné prodloužení všech hlásek v pomalejším úseku. Některé hlásky jsou prodlužovány více než jiné. Proto je nutné, aby databáze umožňovala manipulaci i na úrovni hlásek, nikoliv jen na úrovni segmentů. Úpravy v potřebných temporálních či jiných parametrech se tedy mohou dít i na úrovni hlásek, nejen celých segmentů.

Zde je ještě vhodné uvést určitou návaznost formantové a konkatenační syntézy. Ačkoliv konkatenační syntéza vybízí k tomu ukládat do databáze zvukové záznamy (ať už komprimované či nikoliv), není to jediná možnost. Lze totiž použít i parametrické reprezentace zvuků podobným způsobem, jak to praktikuje formantová syntéza (Tatham & Morton, 2005, str. 31). Jak Tatham & Morton uvádí, tak tyto dvě syntézy se liší především

tím, že jedna odkazuje k abstraktním cílům a druhá ke skutečnému záznamu. Použití parametrů přináší některá ulehčení pro další práci s daty. Mimo jiné je to například menší potřebný úložný prostor (který je pro mobilní zařízení stále nedostatkovým zbožím) a lepší možnost manipulace signálu. Upravovat parametry je vždy snazší cesta než manipulovat přímo se zvukovým signálem. Použití parametrů může nést významnou roli v případě aplikace do specifického prostředí. Výpočetního výkonu není v mobilních zařízeních nikdy dost, stejně tak jako volného prostoru a životnosti baterie. U mobilních řešení stále platí, že jednoduchost a nenáročnost vyhrává nad komplikovaným a výpočetně náročným řešením, byť by přinášelo lepší výsledky (percepčně). Podobně i nepraktičnost vytváření nové databáze pro každý typ afektované mluvy zvlášť by velela spíš k využití parametrické (a tedy manipulovatelné) syntézy. Pokud se totiž pracuje s databází nahrávek mluvčího, jsou už všechny suprasegmentální jevy dané v nahrávce a při požadavku na syntézu například ospalého mluvčího, je nutné vytvořit novou databázi a nechat mluvčího nahrávky namluvit ospale. Vzhledem k náročnosti vytvoření a zpracování databáze stejně tak jako i obtížné pořízení ospalých nahrávek, je zřejmé, že budování nové databáze pro každou afektovanou mluvu zvlášť je téměř nemožné. Ačkoliv každá možnost přináší výsledky rozdílné kvality, nelze říct, že parametrická syntéza by zněla výrazně méně přirozeně oproti syntéze založené přímo na zvukovém záznamu (Tatham & Morton, 2005, str. 31).

Vraťme se však k tomu, jaké jsou v konkatenační syntéze užívány moduly. Již dříve bylo zmíněno, že jeden segment z databáze s druhým je zapotřebí vhodně spojit. Jinými slovy je nutno najít vhodný způsob přechodu mezi koncem jednoho a začátkem druhého segmentu. Při procesu vytváření databáze se za doporučenou praxi považuje vždy značit hranice v místě nejmenší změny. To ovšem nemusí být vždy možné. Nebo i přes nejvyšší snahu se ukáže, že navazující segment má tak odlišné charakteristiky, že při napojení dochází k chybám. Tento problém je řešen tzv. vyrovnáním. V praxi jde o vyhlazení přechodových oblastí, a to nejčastěji v rovině amplitudy. Ideálně by k vyhlazování mělo docházet v oblasti spektra, což

je ale v případě systému založených na principu zvukové nahrávky buď velmi problematické, nebo dokonce nemožné (Tatham & Morton, 2005, str. 33). Co vyhlazování velmi ulehčuje, je parametrické zastoupení zvukového signálu – pak je možné každý parametr systematicky upravit tak, aby byly přechody mezi segmenty co nejplynulejší. Tím je opět přičteno jedno plus k využití parametrické syntézy. Celkově se tím jen opět potvrzuje nezanedbatelná výhoda možnosti manipulace se signálem. Na stranu druhou však jde vždy o to, co je menší zlo – je nutné zvážit i míru snížení celkové přirozenosti.

Ukazuje se, že spojení segmentů na úrovni slabik je výrazně snazší než u segmentů kratších. I přes všechnu snahu mít co nejucelenější databázi s ideálním rozdělením segmentů a efektivním modulem pro usnadnění přechodu mezi segmenty, je zřejmé, že nelze vždy nalézt dokonalou shodu dvou segmentů. Konkatenační systémy proto obsahují i doplňkový modul pro vyhodnocení tzv. ceny pro spojení segmentů. Ten lze jednoduše popsat následovně: pro každé spojení segmentů si klademe otázku - kolik nás bude tato kombinace stát? Není možné nalézt „levnější“ kombinaci? Ve skutečnosti se sčítají dvě ceny. Jednak je to cena výběru jednotky z databáze, neboli jak podobná je zvolená jednotka té cílové ve smyslu zasazení v promluvě, a dále je to cena za spojení jednotek, neboli jak snadno půjde segmenty napojit. Jednoduchým porovnáním pak zjistíme, jestli v databázi není vhodnější kandidát či kandidáti pro zřetězení.

2.5. TTS

TTS (text-to-speech), neboli převod textu na mluvené slovo, je v nejobecnějším popisu systém schopný z psaného slova generovat slovo mluvené. Na první pohled to pro člověka neznalého věci může znít jako velmi snadný úkol. Mnoho lidí může znát TTS systémy buď ze svého počítače, ve kterých je dnes v rámci moderních operačních systémů TTS standardem, nebo z mobilního zařízení, kde je například možnost automatického přečtení SMS zprávy při používání náhlavní soupravy. Jakmile se ovšem začneme zabývat tím, aby generovaný hlas nezněl strojově, dostáváme se do naprosto jiné dimenze. Nestačí jen prosté omezení se na

jednoduchý předpoklad, že tečka znamená intonační pokles, čárka jisté oddělení segmentů a otazník intonační vzestup. A to zde nejde jen o problematičtější interpretaci otazníku jako intonačního vodítka, ale především o celý syntakticko-sémantický rámec věty či promluvy. Pro lepší přirozenost proto musí být v TTS systému kvalitně navrhnutý prosodický modul.

Vzhledem k celkové náročnosti zpracování textu a též zamýšlenému účelu syntézy je nasnadě otázka, zdali stačí pouze syntéza, která bude sice čitelná jako počítačová syntéza, ale její informativní funkce bude jasně zřetelná, nebo na druhou stranu budeme cílit na co nejvěrnější stylizaci do toho, jak by promluva zněla od živého mluvčího. Druhý případ je jasně náročnější a krom jiného by zřejmě vyžadoval i hlubší sémantickou analýzu situace, případně doplnění o nějaký pragmatický modul.

2.6. Fokus vývoje syntetizérů

V prvních krocích umělého tvoření lidského hlasu bylo nejdůležitější měřítko srozumitelnosti (Tatham & Morton, 2005, str. 1). Nejvíce byly testovány jednotlivé hláskové segmenty, které pak tvořily slova. Dobrý systém byl tedy takový, který měl srozumitelné podání na segmentální úrovni. Dnes je stále více kladen důraz nejen na srozumitelnost, ale hlavně na přirozenost. Nejde jen o to text srozumitelně produkovat, ale produkovat ho přirozeně. Další krok po úspěšném sestavení přirozeně znějícího syntetizéru určeného pro produkci řeči za běžných podmínek je kompilace takového systému, který bude schopen produkovat expresívní projevy. Úskalím produkce lidské řeči stroji je to, že jedna, byť i krátká, promluva, kolikrát je vyřčena, tolik bude mít konkrétních realizací na úrovni expresivnosti.

Právě na suprasegmentální úrovni je široce otevřené pole pro další zdokonalování. Zjednodušeně by šlo shrnout, že jde o zdokonalení na dvou úrovních: statické a dynamické. Statickou úrovní je zde myšleno samotné modelování jednotlivých afektivních stavů, jako je zloba, klid, napjatost a další. V tomto směru by byla potřeba se zaměřit na zdokumentování charakteristik jednotlivých typů afektované řeči a následná projekce těchto charakteristik do

parametrů syntetizátoru. Dynamickou úrovní je zde myšleno přiblížení se skutečné přirozené produkci řeči v tom smyslu, že běžně člověk v rámci promluvového úseku, nebo i v rámci celé promluvy, může měnit suprasegmentální charakteristiky, jako jsou například tempo (změna tempa jako upozornění na místo, které je důležité), ladění hlasu (inter- a intra-personální naladění) a mnohé další. Je pouze přirozené, že v průběhu promluvy člověk své mluvní styly mění, a tak je na místě, aby i syntetizér tyto změny uměl kopírovat a zahrnul je do produkce.

3. Znělost

Znělost je důsledek činnosti hlásek, kterou lze definovat jako vlastnost vycházející z přítomnosti základního hlasivkového tónu. V praxi se jedná o distinktivní rys, který od sebe rozlišuje minimální znělostní páry. V nejjednodušším měřítku tak lze od sebe rozlišit párové hlásky, jako jsou například [t], [s] a [t͡s] od [d], [z] a [d͡z]. Konkrétní příklady zkoumaných hlásek a jejich podrobnější popis bude blíže popsán v dalších kapitolách této práce. Nyní se však blíže podívejme na vznik znělosti a popis fonačního ústrojí.

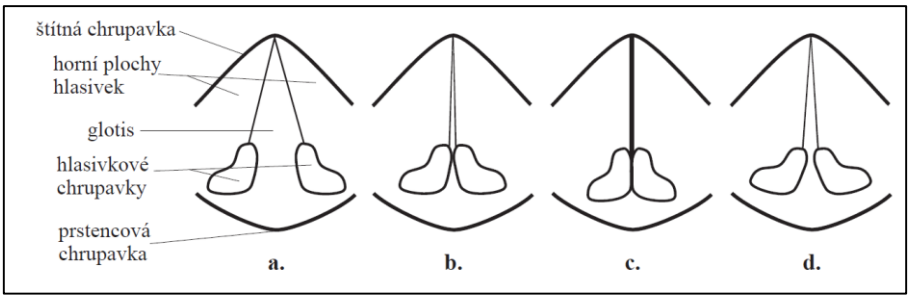
3.1. Hlasivky

Jakkoliv termín základní hlasivkový tón může vzbuzovat představu jednoduchosti, práce hlasivek je ve skutečnosti velmi složitá. Poměrně komplikovaný je i popis hlasivek. Začněme tedy od toho jednoduššího. Velmi častou představou mezi laickou veřejností může být, že hlasivky fungují jako nezávislý orgán, který vlastním úsilím produkuje hlas. Často se tato představa spojí ještě s tím, že jejich činnost je zřejmě svázána s výdechem. Hlas je pak dále v puse proměněn na řeč. Tyto laické představy se na dalších řádcích pokusí tato práce opravit a upřesnit.

Pro lepší představu toho, jak lze kmitáním vytvořit zvuk, můžeme použít pouťový balónek. Naplníme-li balónek vzduchem, který pak budeme vypouštět sevřeným hrdlem, vytvoříme tón. Na podobném principu fungují i hlasivky. Ti, kteří dávali ve škole pozor, si mohou pamatovat, že hlasivky kmitají. Otázka je, jak to dělají a proč. Začněme otázkou proč. Zvuk ve skutečnosti není nic jiného než kmitajícím tělesem vytvořené vlnění, které pak receptory v našem uchu vnímají jako určitý signál a mozek vyhodnotí jako zvuk. V našem případě to jsou právě hlasivky, které kmitají. Daleko zajímavější je však otázka, jak kmitání hlasivky vlastně dosáhnou. Zde je nutné upozornit na to, že užití činného rodu v předešlé větě není zcela přiléhavé. Hlasivky totiž aktivně kmitavé pohyby nevytvářejí, nýbrž jen pasivně kmitají díky vzduchu proudícímu z plic.

Hlasivky mají schopnost zcela uzavřít cestu unikajícímu vzduchu z plic. To je dáno jednou z funkcí hlasivek, kdy mají za úkol pomoci zpevnit celkové svalové napětí v těle, zvláště pak v oblasti břišních svalů, a umožnit tak lepší efektivitu práce. Vzepření se se zadrženým dechem nám dává možnost vyvinout větší sílu. Až později ve vývoji člověka se k této funkci přidala schopnost produkovat kvazi-řečový a později i řečový signál.

Krom úplného zavření či otevření cesty z plic se umí hlasivky uzavřít částečně. To je klíčový moment pro většinu znělých hlásek. V této fázi začínají hlasivky při výdechu kmitat a vytvářet tak základní hlasivkový tón. Jak ale bylo řečeno už dříve, je to poměrně komplikovaný proces. Při bližším zkoumání hlasivek totiž narazíme na dva zajímavé faktory. Jednak nejsou hlasivky uspořádány kruhově kolem dýchací trubice (jak bylo zjednodušeně demonstrováno na příkladu s balónkem), nýbrž jsou na jedné straně jakoby sepnuty a rozevírají se na straně druhé. Tím vzniká jakýsi trojúhelníkový prostor kolmý na průdušnici. Pohyblivá (z hlediska uzavírání a otevírání) je pouze část otevřená (základna trojúhelníku), která se umí přiblížit anebo oddálit. Pohyb pouze jedné části hlasivek je z pohledu ergonomie výhodnější. Pro lepší představu viz obrázek (Obrázek 3).



Obrázek 3 Schematický nákres základních poloh hlasivek (shora): **a.** dýchání, **b.** fonační postavení, **c.** ráz, **d.** šepot (Skarnitzl, Znělostní kontrast nejen v češtině, 2011, str. 30).

Trojúhelníkový charakter hlasivkové štěrbiny však není jediná zajímavost vhodná popisu. Hlasivky totiž nejsou tvořeny tenkou jednolitou blankou (jak bylo opět zjednodušeně demonstrováno na příkladu s balónkem), ale jsou tvořeny z horní a spodní části (bráno řezem podél vertikální osy). Obě části se pohybují na sobě relativně nezávisle. Je zřejmé, že pokud začneme odkrývat detailní akustické děje probíhající v hlasivkách, tak nám i takto jednoduchý popis dává vědět, že to bude odkrývání poměrně komplikované.

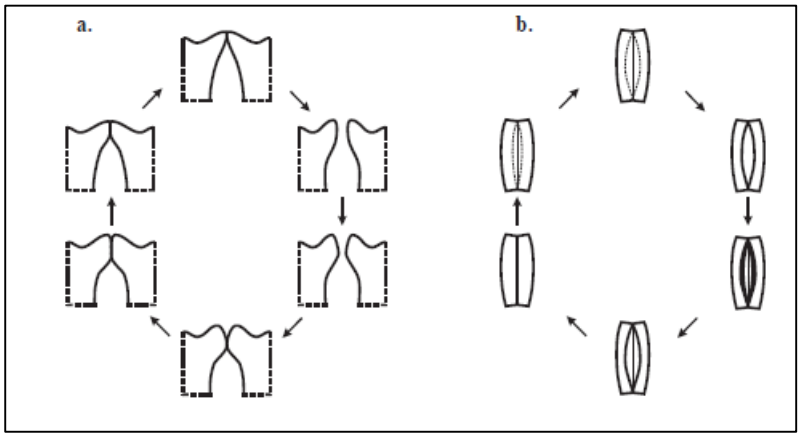
3.2. Popis kmitání hlasivek

Polohu hlasivek lze rozdělit na čtyři základní stavy. Prvně je to stav, kdy volně dýcháme, a hlasivky jsou v klidové otevřené pozici. Dále je to pozice, kdy jsou hlasivky ve fonačním postavení. To je situace, kdy jsou hlasivky sice uzavřeny, ale ne pevně sevřeny. Začne-li z plic proudit výdechový proud vzduchu, hlasivky se rozkmitají a začnou fonovat. Třetí pozice je stav, kdy jsou hlasivky pevně sevřeny a průchod vzduchu vůbec neumožňují. Tohoto stavu se využívá jednak jako ochrana před vniknutím potravy do dýchacích cest, dále v situacích, kdy je zapotřebí vyvinout vyšší napětí svalové soustavy člověka (jak bylo popsáno dříve), a dále pak v situaci, kdy se při řeči vytváří tak zvaný hlasivkový ráz. Poslední pozicí je nastavení hlasivek pro šepot, kdy dochází k částečnému přiblížení hlasivek, ale stále je umožněno proudění velkého množství vzduchu ve srovnání se základním fonačním postavením (Skarnitzl, Znělostní kontrast nejen v češtině, 2011, str. 30).

3.2.1. Fáze činnosti hlasivek

Vlastní kmitání hlasivek lze rozdělit do několika fází. Nejprve je nutné, aby hlasivky zaujaly fonační postavení. Jakmile je ho dosaženo, začne se pod hlasivkami díky výdechovému proudu vzduchu tvořit tlak, který postupně narůstá. Tento tlak nazýváme subglotální (p_{sg}). S jeho postupným zvětšováním dojde nejprve k oddálení spodní a pak horní části hlasivek. Jakmile se tak stane, rozraženými hlasivkami začne proudit nahromaděný proud vzduchu. Jeho rychlým unikáním dojde ke zředění vzduchu na místě, kde dříve byly hlasivky semknuté. Dle Bernoulliho aerodynamických zákonů v tomto místě následně dojde k podtlaku, který k sobě přitáhne spodní a následně i horní část hlasivek (Skarnitzl, Znělostní kontrast nejen v češtině, 2011, str. 31). Soubor všech těchto fází tvoří cyklus, který se po dobu fonace kvaziperiodicky opakuje (viz Obrázek 4). Důsledkem fonační činnosti je vypouštění vln zředěného a zhuštěného vzduchu, který následně prochází nadhrtanovými dutinami a umožňuje nám tak formovat zvuky a následně řeč.

Do procesu fonace se krom Bernoulliho zákonů zapojují i další fyzikální jevy. Jedním z nich je van den Bergova myoelastická-aerodynamická teorie tvorby hlasu. Van den Berg uvádí, že základní frekvence kmitání hlasivek je dána pěti na sobě nezávislými faktory: efektivní hmotou hlasivek, napětím hlasivek, plochou průřezu hlasivkové štěrbiny, hodnotou subglotálního tlaku a tlumením hlasivek (Skarnitzl, Znělostní kontrast nejen v češtině, 2011, str. 32). Zjednodušeně lze vyjít z názvu teorie a rozložit si ji na dva aspekty. První aspekt, myoelasticita, souvisí s nervovým a svalovým ovládáním elasticity hlasivek a měkkých struktur hrtanu. Druhý aspekt, aerodynamicita, souvisí s prouděním vzduchu přes hlasivky a tlaků potřebných k rozkmitání hlasivek. Další výzkumy v této oblasti však ukázaly, že výpočet potřebného subglotálního tlaku k iniciaci fonace je ještě o něco komplikovanější.



Obrázek 4 Schematický nákres hlasivkového cyklu: a. pohled zepředu, b. pohled shora **(Skarnitzl, Znělostní kontrast nejen v češtině, 2011, str. 31)**.

Významný výzkum v činnosti hlasivek shrnul Jiang a kolektiv. Počátkem tohoto století postuloval tři dle jeho zkoumání nejdůležitější faktory, které způsobují kmitání hlasivek. Prvním z nich je již výše zmiňovaný Bernoulliho zákon. Druhým faktorem je již také zmíněná elasticita hlasivek, která způsobuje samovolné navracení hlasivek do neutrálního stavu, tedy stavu před oddálením. Posledním faktorem, který Jiang zmiňuje, je důsledek snižování subglotálního tlaku po uvolnění hlasivek v rámci jednoho cyklu. Po uvolnění nahromaděného tlaku dochází bezpodmínečně k jeho snížení. To dává za příčinu snížení sil působících na hlasivky, pro kteréžto je poté snazší se navrátit to původní polohy a průchod vzduchu opětovně uzavřít. Těmito třemi faktory tedy poměrně výstižně Jiang a kolektiv shrnuje dosavadní výzkumy.

V tuto chvíli zde stojí za opětovné zdůraznění to, že hlasivky samy se na kmitání aktivně nepodílejí. Svalovou aktivitu vyvíjejí pouze pro prvotní semknutí, ale po celou dobu fonačního cyklu již nadále zůstávají svalová vlákna oblasti hlasivek bez aktivity, alespoň co se týče jejich podílu na kmitání. Rozsah kmitání hlasivek při běžném hovoru je přibližně v rozmezí 80 Hz až 300 Hz, tedy 80 až 300 kmitů za vteřinu. U mužů se průměrný rozsah uvádí 100-150 Hz, zatímco u žen se uvádí rozsah 200-300 Hz (Palková, 1994, str. 57). Tomuto kmitání hlasivek, které tvoří základ hlasu, se říká základní tón. Pro porovnání lze uvést, že můry kmitají křídly v rozsahu 150 až 200 kmitů za vteřinu, včely a mouchy již výrazně přesahují počet 200 kmitů za vteřinu a komáři dosahují hodnoty 500 kmitů za vteřinu (Biewener, 2003, str. 140). Přičemž obecně platí, že čím vyšší frekvence kmitání, tím vyšší tón vnímáme. Jednotlivé impulzy pak jako souvislý tón začínáme vnímat přibližně od 20 kmitů za vteřinu.

Vraťme se však k fyzikálnímu popisu kmitání hlasivek. Zajímavý výzkum představil na přelomu osmdesátých a devadesátých let minulého století Titze. Ten ve své teorii pracuje se dvěma prvky. Prvním z nich jsou již představené Bernoulliho aerodynamické zákony, které doplňuje o interakce s vokálním traktem jak pod tak nad hlasivkami. Druhým a tím podstatnějším prvkem je opožděná změna supraglotálního tlaku vzduchu ve vokálním traktu oproti pohybu hlasivek (Skarnitzl, Znělostní kontrast nejen v češtině, 2011, str. 35). Ve zkratce poukázal na to, že unikající proud vzduchu nad hlasivkami ve fázi, kdy se hlasivková štěrbina zpět uzavírá, pomůže hlasivkám se uzavřít. To je způsobeno tím, že proudící sloupec vzduchu nad hlasivkami má i při uzavírání hlasivek ještě určitou hybnou energii, která jakoby vcucne hlasivky a uzavře je. Titze na matematických modelech prokázal, že bez zmíněného druhého prvku by ke kmitání nedocházelo.

Vezmeme-li jako fakt, že se hlasivky na kmitání aktivně svou svalovou činností nepodílí, musí v systému existovat prvek, který bude hlasivkám energii pro rovnoměrné kvaziperiodické kmitání dodávat. Hirano se tedy zamýšlel nad tím, kde se bere ona síla, která

kmitání hlasivek udržuje. Základní kámen jeho úvah představuje zákon o zachování energie. Způsobů jak vysvětlit tento elementární zákon je několik, my si však vystačíme se zjednodušeným vysvětlením, že množství energie v rámci soustavy zůstává stejné, pouze se může přeměnit v jiné formy energie. Obecně vzato nelze sestavit perpetuum mobile, neb jsme zatím nebyli schopni sestavit stroj se stoprocentní účinností. Velmi laicky řečeno, jakákoliv soustava energii jednak přijímá, pak zpracuje a nakonec energii ve formě nějakého výsledku vydá. Snadno si to lze představit například na žárovce. Do žárovky vstupuje energie ve formě elektřiny. V žárovce dojde k několika procesům, z nichž ten kýžený je pro nás světlo. Ve skutečnosti je však, procentuálně vzato, světlo u klasických žárovek odpadovým sekundárním jevem, neboť jen asi 10% energie se v běžné žárovce přemění ve světlo. 90% účtu za vysvícenou elektřinu ve skutečnosti platíme za teplo, které žárovka vygeneruje. Dalo by se tedy říct, že v žárovce dochází k úniku energie. Ve skutečnosti je ale pouze přeměněna na teplo. Celkové množství energie tedy zůstává neměnné, pouze její forma se mění (Účinnost (fyzika), 2013).

Pojďme se ještě podívat na druhý velmi jednoduchý příklad. V prvním příkladů šlo o energii ve formě elektřiny, zde bude příklad demonstrován v podobě mechanické energie, která je pro přiblížení činnosti hlasivek bližší. V předchozím příkladu jsme si ukázali, jak je vložená energie rozdělena na několik odchozích složek. Tak se tomu děje prakticky při každém přenosu energie. Cestujeme-li na bicyklu, je to stejné. Naše svaly vyvíjí určitou energii, kterou pomocí mechanického soukolí převádíme do hnací síly kola. Energie, kterou svaly vyvineme, však není stoprocentně přeměněna na pohybovou energii. Nežli se rozjedeme, je nutné překonat několik sil, které jdou proti nám. Jednou z nich je mechanické soukolí bicyklu, kde dochází ke ztrátě energie díky tření. Pro zjednodušenou představu si lze představit zarezlé soukolí, kde nejprve musíme odtrhnout přirezlé části kovu od sebe. To je samozřejmě extrémní příklad, ale pokud kolo není namazané, dochází zde stále ke znatelnému tření kovu o kov. Mazivo toto tření výrazně eliminuje a efektivně tak snižuje energii

vydávanou na nežádoucí činnosti. Ne nadarmo se tedy říká „Kdo maže, ten jede“. Cílem těchto dvou příkladů bylo znázornit, že do soustavy je nutno dodávat více energie, než jen kolik z ní chceme v určité podobě energie odčerpat.

Faktor, který se velkou měrou podílí na způsobu kmitání hlasivek, je skladba jejich tkáně. Hlasivky jsou tvořeny jednak obalovou tkání a pak vnitřní svalovou tkání. Obal je tvořen epitelem a prostřední vrstvou. Vnitřní část, tělo, je pak tvořeno hlubokou vrstvou a svalem. Z hlediska kmitání je důležité, že hlasivky nekmitají jako celek. Horní a spodní část kmitají s jinou fází, k čemuž navíc se obal hlasivek může pohybovat do jisté míry nezávisle na vlastním tělu.

3.3. Fonační hlasový práh

Podmínky, které je nutné splnit pro zahájení a udržení fonace, se dnes definují jako následující čtyři: (1) dostatečný rozdíl mezi subglotálním a supraglotálním tlakem, (2) dostatečně silný expirační proud vzduchu, (3) dostatečné přiblížení hlasivek a (4) dostatečně nízký glotální odpor (Skarnitzl, Znělostní kontrast nejen v češtině, 2011, str. 39). Jedním termínem se toto popisuje jako fonační prahový tlak (p_{pr}). Touto hodnotou se definuje minimální hodnota subglotálního tlaku a jako taková popisuje snadnost fonace. Pro svou praktičnost je fonační prahový tlak hojně využíván při diagnostice hlasivek a jejich zdravotního stavu.

Z přímých vztahů dále fonační prahový tlak ovlivňuje tlumení hlasivkových tkání, rychlost slizniční vlny a šířka glottis před fonací. Z toho vyplívá, že čím užší je tlakové hrdlo před fonací, tím nižší je práh pro fonaci. Tento vztah však není čistě lineární, jak by se na první pohled mohlo zdát. U hodnot nižších než 1 mm dochází k nelineárním projevům a pod tuto hranici již vztah linearity neplatí (Skarnitzl, Znělostní kontrast nejen v češtině, 2011, str. 40). Rovněž platí, že čím větší tloušťka hlasivek, tím snazší zahájení a udržení fonace. Nepřímo dochází k ovlivnění p_{pr} i parametry jako viskoelastivita a hydratace, kdy platí, že dehydratované hlasivky mají práh vyšší, zatímco optimálně hydratované hlasivky nižší.

Posledním poměrně zajímavým faktem je to, že udržet hlasivky ve fonaci je snazší, než je z nuly rozkmitat. Práh pro nástup fonace bývá zpravidla vyšší než práh pro její odeznění. Je to ve své podstatě stejný princip jako na kole nebo v automobilu, kdy rozjet se je energeticky náročnější než rychlost jen udržovat.

3.4. Fyziologické ovládání znělosti

Jak bylo již nastíněno v oddílech 3.1 a 3.2.1, hlasivky se při fonaci aktivně pouze přiblíží do fonačního postavení (addukce) a vlastní fonace pak probíhá bez jejich aktivního zapojování. Konkrétně se zapojí dva svaly, a to laterální kriko-arytenoidní sval a inter-arytenoidní sval (Skarnitzl, Znělostní kontrast nejen v češtině, 2011, str. 44). Hlasivky následně pouze udržují své nastavení. Toto platí pro hlásky znělé. Při střídání hlásek znělých a neznělých, což je situace, která nastává při běžné promluvě, je to již poněkud komplikovanější. Pro snazší pochopení činnosti hlasivek je nutné velmi zjednodušeně popsat princip svalové součinnosti.

3.4.1. Koordinovaná svalová činnost

Každý pohyb vyžaduje dokonalou souhru mezi všemi zapojenými svaly a svalovými skupinami. U zdravého člověka platí, že pro zajištění hladkého průběhu pohybu je nutné, aby se při kontrakci jednoho svalu skupiny zároveň tlumil tonus jeho antagonisty. Tomuto mechanismu se říká reciproční inhibice (Troja, Druga, & Pfeiffer, 1990, str. 34). Zjednodušeně řečeno to znamená, že pokud je zapotřebí k pohybu A jednoho svalu a k pohybu B v opačném směru svalu druhého, pak se svalové zapojení při změně pohybu střídá postupně, nikoliv skokově. Svalovou součinnost Druga a Pfeiffer popisuje na svalstvu inervovaném z motorických jader míšních nervů, což jsou například bicepsy a tricepsy (Linc & Doubková, 2003, stránky 129, 131), ale činnost hlasivek je v tomto ohledu analogická. Reflexní oblouk mozkových nervů, ze kterého jsou hlasivky inervované, je funkčně stejný jako reflexní oblouk míšní (Králíček, 2004, str. 136).

Za oddalování (abdukci) hlasivek a tím i přerušení fonace je zodpovědný posteriorní kriko-arytenoidní sval. Při produkci střídavě znělých a neznělých hlásek se za použití elektromyografu ukázalo, že posteriorní kriko-arytenoidní sval (abduktor) se postupně aktivuje a inter-arytenoidní sval (adduktor) se postupně deaktivuje (Skarnitzl, Znělostní kontrast nejen v češtině, 2011, str. 44).

3.5. Filtrová teorie produkce řeči

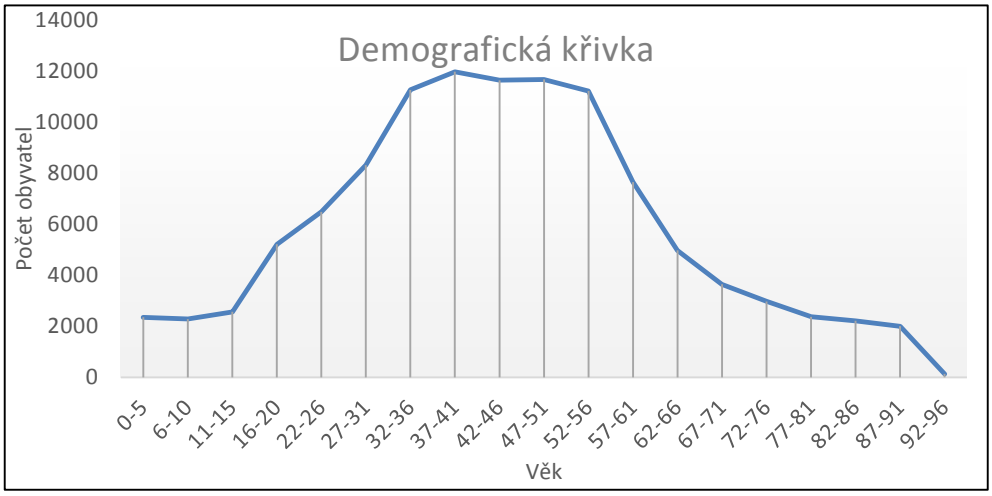
To, jaká je výsledná znělostní charakteristika, je dobře popisováno filtrovou teorií produkce řeči. Jak bude z dalšího popisu zřejmé, největší uplatnění najde při popisu produkce vokálů.

V minulosti se mezi fonetiky vedly diskuse, jaký zvuk vlastně hlasivky vydávají. Dlouho se mělo za to, že je to zvuk tichý a řezavý. Dle všeho je to ale chybná představa. Filtrová teorie produkce řeči totiž chápe nadhrtanové dutiny jako zvukový filtr, který některá frekvenční pásma tlumí a jiná nikoliv. O jaká pásma se konkrétně jedná, je dáno především nastavením vokálního traktu pro každou hlásku. Utlumeným frekvenčním pásmům se říká formanty. Důvod, proč základní hlasivkový tón musí být silný, je především ten, že vokální trakt je pro šíření zvuku ve skutečnosti velmi nepříznivé prostředí. Jeho měkké, vlhké a členité prostředí funguje jako silný tlumící prvek. Rovněž z podstaty věci není schopen do zvukového signálu přidávat frekvence nové, pouze ty stávající tlumit. Z toho vyplývá, že základní hlasivkový tón musí obsahovat celou škálu frekvencí, pouze jsou po cestě některé z nich tlumeny více a jiné méně.

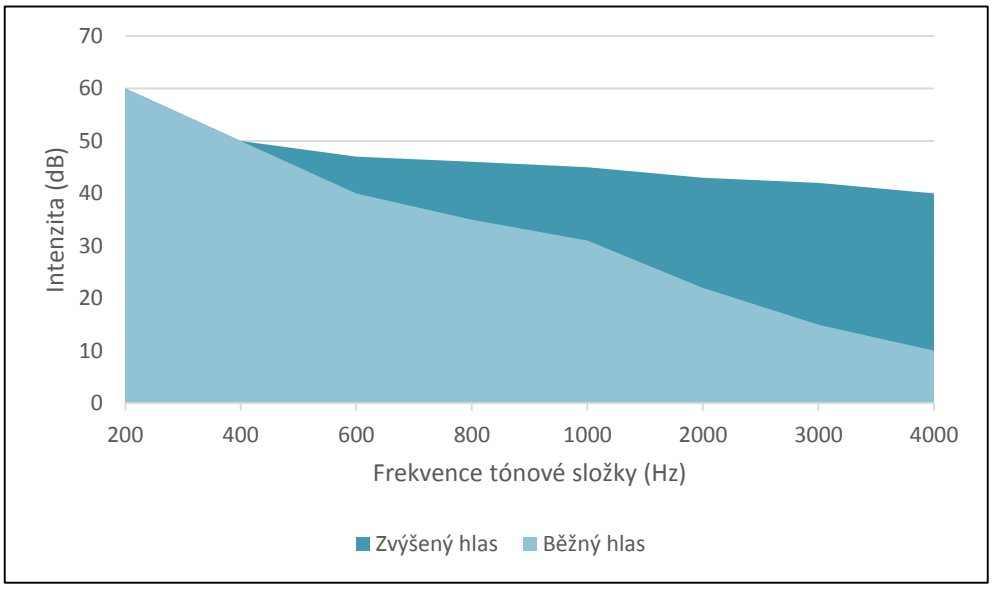
Další veličinou, která nám přibližuje kvalitu hlasivkového tónu, je spektrální sklon. Spektrální sklon je veličina, která popisuje, jaká je spektrální obálka daného zvukového signálu. Představit si ji lze jako demografickou křivku obyvatelstva. Na té je znázorněno, jaký je počet obyvatel v daných věkových kategoriích a odvodit si, zdali je v populaci převaha mladších, středněvěkých, či starších obyvatel (Obrázek 5). Namísto věkových kategorií se ve spektrálním sklonu počítají frekvenční složky. Jsou znázorněny od těch nejnižších po ty

nejvyšší a místo počtu osob se znázorňuje dominantnost daných frekvencí (jinými slovy jejich amplituda). Tím nám vznikne obrázek toho, jestli v tónu převažují frekvence vysoké či nízké. Konkrétní průběh demografické křivky a spektrální obálky řeči je samozřejmě odlišný.

Při běžné produkci řeči má výsledný řečový signál spektrální sklon -6dB na oktávu. To znamená mírný pokles v oblasti vyšších frekvencí. Při přechodu mezi vokálním traktem a okolním prostředím však dochází v důsledku radiační impedance ke změně spektrálního sklonu. Efektivně na rtech dochází ke změně o +6dB na oktávu, z čehož nám vychází, že hlasivkový signál uvnitř vokálního traktu má spektrální sklon -12dB na oktávu. To nám naznačuje, že v tónu hlasivek je znatelná převaha nižších frekvencí nad vysokými. V praxi to pak pro naše ucho představuje tón podobný více hlubokému brumu nežli řezavému vysokému tónu. Jako zajímavost lze



Obrázek 6 Příklad demografické křivky obyvatelstva.



Obrázek 5 Ukázka spektrálního sklonu lidského hlasu. Při vyšší hlasitosti se začíná sklon vyrovnávat, což je způsobeno větším zastoupením frekvencí ve vyšších frekvencích. Pro lepší názornost byly zvýrazněny rozdíly v průběhu obou křivek a tak nemusí zcela odpovídat realitě.

ještě zmínit fakt, že při hlasité řeči se spektrální sklon vyrovnává až na úroveň -3dB na oktávu. Pro představu spektrálního sklonu a vlivu hlasitosti na jeho průběh viz obrázek (Obrázek 6). Spektrální sklon jako takový číselně popisuje jednu z kvalit hlasu a lze tak použít jako diagnostickou pomůcku pro měření fonačních modulací.

4. Konsonantická a vokalická artikulace

Produkce řeči pochopitelně nestojí pouze na činnosti hlasivek. Důležitou částí tvorby řeči je artikulace, což je proces v nadhrtanových dutinách, do kterého se zapojují artikulační orgány. Mezi ty se řadí rty, zuby, dásňový výstupek, tvrdé patro, měkké patro, čípek, zadní stěna hrdelní dutiny, spodní čelist a jazyk (Palková, 1994, str. 64). Bližší popis si ponechejme až pro konkrétní popis tvorby exploziv dále v této práci, prozatím se spokojme pouze s tímto prostým výčtem.

Základní dělení artikulace dle činnosti jednotlivých artikulačních orgánů se dělí na artikulaci vokalickou a konsonantickou (Palková, 1994, str. 68). Pokud se artikulační orgány nastaví tak, že výsledkem je zvuk s převahou tónové složky, bavíme se o artikulaci vokalické. Principiálně lze říct, že při vokalické artikulaci pouze měníme velikost a tvar dutiny ústní, nedochází však k tvorbě výrazných překážek. K těm dochází při artikulaci konsonantické, při které je cílem vytvořit unikajícímu proudu vzduchu překážku ve formě výrazného zúžení či úplného přepažení. Výsledkem je zvuk, ve kterém má převahu šumová složka. Pro určení toho, která složka ve zvukovém signálu převažuje, lze použít více metod. Těmi se však v tuto chvíli nezabývejme a pouze pro laickou představu si můžeme nahlas vyslovit slovo „*zase*“. Se znalostmi ze základní školy jsme schopni v tomto slově identifikovat vokály a konsonanty. Řeknete-li toto slovo pozorně a pomalu a zaposloucháte-li se do jeho zvuku, můžete si udělat představu o tom, co je to tónová a šumová složka. Hlásky [z] a [s] obsahují převahu šumové složky a hlásky [a] a [ɛ] obsahují převahu tónové složky. Za zmínku ještě stojí fakt, že dělení hlásek na vokály a konsonanty dle jejich zvukové kvality není až tak jasně bipolární, jak by se na první pohled mohlo zdát. V češtině patří k problematičtějším například hláska [j], která je velmi podobná hlásce [i].

Tato práce se zaměřuje především na tvorbu exploziv, což jsou hlásky spadající do skupiny konsonantů. Ponechme tedy bližší popis vokalické artikulace stranou a podívejme se

rovnou na artikulaci konsonantickou. Jak už bylo řečeno dříve, hlavním prvkem konsonantické artikulace je vytvoření překážky výdechovému proudu vzduchu někde v nadhrtanových dutinách. Jedním ze dvou definujících prvků konsonantické artikulace je právě místo oné překážky. Mluvíme pak o místě artikulace. Druhým definujícím prvkem je způsob, jakým je překážka vytvořena. Mluvíme pak o způsobu artikulace. Místo artikulace je dáno artikulačním orgánem, který se na artikulaci podílí. Jejich výčet byl uveden v úvodu této kapitoly. Konkrétní příklady pak budou uvedeny dále. Způsobem artikulace se konečně dostáváme blíže k explozivám.

4.1. Explozivy

Explozivy, česky také hlásky závěrové, jsou hlásky, při jejichž artikulaci dochází k úplnému uzavření cesty výdechovému proudu vzduchu (Palková, 1994, str. 74). V určitém okamžiku je pak cesta výdechovému proudu vzduchu naráz otevřena, čímž vzniká velmi charakteristický zvuk – exploze. Ta je dána rychlým vypuštěním vzduchu, který se nahromadil před překážkou. Jelikož se tato práce zabývá českými explozivami, bude-li řeč o explozivách, budou vždy míněny explozivy české. To samé platí i o ostatních hláskách.

V češtině rozeznáváme několik exploziv. Uvedeny budou postupně dle místa jejich tvoření od rtů dále směrem do hloubi nadhrtanových dutin. První uvedená hláska je vždy neznělá, u druhé hlasivky kmitají a je tedy znělá. Nejprve jmenujme explozivy, které jsou tvořeny semknutím horního a dolního rtu. Říká se jim bilabiální explozivy, česky také obouretné, alternativě ještě retoretné. Patří mezi ně [p] a [b]. Někteří autoři mezi ně ještě zahrnují nazální hlásku [m] (Palková, 1994, str. 223). V této práci však nebudeme s kategorií nazálních exploziv pracovat. Další sadou českých exploziv jsou hlásky [t] a [d]. Dle místa jejich tvoření se jim říká alveolární explozivy, česky též dásňové přední. Artikulovány jsou vytvořením překážky v místě alveol, tedy dásňového výstupku těsně za zuby. Překážka je tvořena špičkou jazyka opírající se o alveoly. Následující sada exploziv je tvořena na tvrdém

patře, odborně též palatu. Odtud plyne jejich název, explozivy palatální, česky též tvrdopatrové. Závěr je tvořen opět pomocí jazyka, v tomto případě pomocí hřbetu jazyka. Ten se přitiskne k tvrdému patru, čímž je vytvořena žádaná překážka. Explozivy, které jsou takto tvořeny, jsou [c] a [ɟ]. Díky problematické syntéze těchto hlásek v programu HLsyn bude tento pár z výzkumu vynechán. Poslední sadou jsou explozivy velární. Ty jsou tvořeny přitisknutím zadní části jazyka k vélu, tedy k měkkému patru. Patří mezi ně [k] a [g]. Tím je uzavřen výčet českých explozivních fonémů.

4.1.1. Specifika artikulace znělých exploziv

Jak již bylo podrobněji vysvětleno dříve, znělost hlásek je dána přítomností základního hlasivkového tónu. Ten je vytvářen pomocí proudu vzduchu unikajícího z plic. V předešlé kapitole jsme se však dozvěděli, že explozivy jsou tvořeny pomocí úplné překážky v nadhrtanových dutinách. Po omezenou dobu je tak výdechovému proudu vzduchu znemožněno unikat. Před překážkou se díky nahromaděnému vzduchu zvýší supraglotální tlak, čímž dojde ke srovnání transglotálního tlaku. Důsledek vyrovnání transglotálního tlaku je nemožnost udržet fonaci (Skarnitzl, Znělostní kontrast nejen v češtině, 2011, str. 51). Na první pohled by se chtělo udělat závěr, že je tedy možné produkovat pouze neznělé explozivy. Z praxe však bezpečně víme, že tomu tak není. Lze to demonstrovat jednoduchou sadou experimentů, které si může čtenář vyzkoušet i sám, bude-li chtít. Nechť má ale čtenář na paměti, že v průběhu jednotlivých experimentů se pracuje s dechem, což může mít za následek motání hlavy v důsledku hyperventilace.

Pro zjednodušení experimentu v tuto chvíli předpokládejme, že místo závěru v nadhrtanových dutinách nemá vliv na délku možné fonace. Pro zachování co nejsnazší demonstrace budeme předpokládat závěr bilabiální. Níže následují instrukce pro jednotlivé experimenty.

(1) Nadechněte se. Po nádechu zkuste říct velmi dlouhou hlásku [m], tak jako byste například chtěli meditovat pomocí známého [ɔːm]. Při tomto experimentu, můžete sledovat například to, že hláska [m] je znělá. Dotknete-li se rukou spodní části ohryzku, můžete cítit jemné vibrace, což je projev aktivních hlasivek, které právě tvoří základní hlasivkový tón. Rovněž si povšimněte toho, že ačkoliv máte zavřená ústa, proud vzduchu z plic přes hlasivky a dále nadhrtanové dutiny stejně uniká. To si můžete ověřit, pokud si dáte konečky prstů pod nos. Ucítíte, jak vám z nosu uniká proud vzduchu. Alternativně to lze ověřit i pomocí zrcátka, které se u nosu zamlží.

(2) Nyní si zkuste, co se stane, pokud si v průběhu bodu (1) zacpete nos. Fonace i nadále po nějakou dobu pokračuje (můžete hlásku [m] slyšet). Patrně jste si však všimli, že vaše tváře se nafoukly a spodní čelist poklesla do nižší polohy. Delší fonace jste dosáhli tím, že se vám zvětšil objem nadhrtanových dutin. Dokud jste ho mohli zvětšovat, mohli jste i fonovat. Ve chvíli, kdy se již nemohou nadhrtanové dutiny zvětšovat, fonace ustane. Tím jsme si ověřili, že bez výdechového proudu vzduchu se hlasivky nerozkmitají.

(3) Při běžné mluvě je jasné, že jen stěží využijeme těch mechanismů, kterých jsme si v předchozích bodech všimli. Musí tudíž existovat i další mechanismy. To si nyní ověříme. Začátek bude stejný jako v bodě (2), ovšem s tím rozdílem, že si dáme pozor na to, aby nám neklesala spodní čelist (mějme zuby stále „zaťaté“) a nenafukovaly se nám tváře (například rukou si je tiskneme a kontrolujeme, že se nám nenafukují). Pokud provádíte postup správně, můžete si všimnout toho, že fonace přetrvává výrazně kratší dobu než v předchozích pokusech. K přerušení fonace dojde v krátkém sledu po zacpání nosu, nicméně stále je zřejmé, že i se všemi výdechovými cestami uzavřenými lze po krátkou dobu fonovat. Mechanismy, které to umožňují, se pokusíme odhalit v dalších bodech.

(4) Znovu proveďte pokus z bodu (3). Postup dodržujte stejný, pouze se zaměřte na to, co dělá jazyk. Nejlépe pozorovatelná je situace ve chvíli, kdy se přerušuje fonace. Všimněte si, že jazyk je v jiné pozici, než když uvolněně dýcháte. Jazyk se posunul jakoby dopředu dolů. Tím se zvětšil prostor nadhrtanových dutin a bylo umožněno fonovat i se semknutými rty a zacpaným nosem.

(5) Mechanismus popsaný v bodě (4) však není jediný. Druhý velmi výrazný mechanismus je pozorovatelný v oblasti hrtanu samotného. Pokud už máte předešlé body dostatečně procvičené, můžete si přestat jednou rukou hlídat tváře a dotknout se prsty krku v oblasti ohryzku. Při provádění experimentu z bodu (3) si můžete všimnout, že se celý ohryzek hýbe směrem dolů. Tím, že se hrtan posune dolů, se zvětší prostor nadhrtanových dutin a je nám umožněno déle fonovat. Tento mechanismus (5) spolu s výše popsaným mechanismem (4) hraje významnou roli ve schopnosti udržet fonaci i po dobu tvoření úplného závěru u exploziv. Krom těchto dvou mechanismů je pro úplnost zapotřebí zmínit ještě zvýšení měkkého patra a posun hltanové stěny dozadu jako dalších mechanismů umožňujících zvětšení nadhrtanových dutin. Ty však nelze takto snadno demonstrovat.

Na to, že musí existovat nějaké kompenzační mechanismy, poprvé poukázal Martin Rothenberg v roce 1968 (Skarnitzl, Znělostní kontrast nejen v češtině, 2011, str. 51). Při

tvorbě bilabiální explozivy mají běžné nadhrtanové dutiny objem přibližně 50 ml. K vyrovnání tlaků pod hlasivkami a nad (a tím pádem k ukončení fonace) dochází již za 4 ms. V praxi by to znamenalo nanejvýše jeden hlasivkový kmit. Významný výzkum v této oblasti přinesl John Westbury v roce 1983 (Skarnitzl, Znělostní kontrast nejen v češtině, 2011, stránky 54-55). Podobně jako i Rothenberg dochází k závěru, že existují aktivní a pasivní metody zvětšení objemu nadhrtanových dutin. Jako pasivní uvádí zvětšování vokálního traktu v důsledku působení zvýšeného tlaku na měkkou tkáň stěn traktu. Aktivní pak popisuje jako takové mechanismy, které jsou způsobeny svalovou činností. Řadí mezi ně mechanismy uvedené výše v bodě (5) pokusů – snížení hrtanu (demonstrováno bodem (5)), vyklenutí měkkého patra vzhůru, posun kořene jazyka dopředu a snížení těla a čepele jazyku ke spodině ústní (tento a předchozí bod byly společně demonstrovány v pokusech v bodě (4).

Užití mechanismů není ve všech případech stejné a často se jedná o kombinaci jen některých mechanismů. Kombinatorikou aktivních a pasivních mechanismů se zabývalo více výzkumných týmů, zmiňme však výzkum Maria A. Svirsky (Svirsky, a další, 1997). Pomocí elektromagnetické artikulografie a měření intraorálního tlaku Svirsky a jeho tým došli k závěru, že posun jazyka je výraznější u [b] než u [p]. Intraorální tlak pak u [p] roste skokovitě, zatímco u [b] je nárůst postupný. To lze vysvětlit potřebou zachovat určitý transglotální tlak u znělého [b] po celou dobu závěru.

Po vysvětlení způsobu udržení fonace při artikulaci exploziv je nasnadě otázka, jaký vliv má na udržení fonace místo artikulace. Již poslední zmíněný výzkum dává tušit, že místo artikulace bude hrát při udržení fonace svou roli. Vezmeme-li v potaz dva zásadní aspekty udržení fonace, a to vyrovnání transglotálního tlaku a velikost poddajné plochy v orálním traktu, dojdeme k závěru, že u přednějších exploziv bude snazší fonaci udržet než u exploziv zadnějších. Jednak u zadních exploziv (například české [g]) dojde k dřívějšímu vyrovnání transglotálního tlaku v důsledku menšího objemu prostoru mezi hlasivkami a závěrem (oproti

předním explozivám jako například u českého [b]), a dále pak udržení fonace znesnadňuje menší plocha stěn vokálního traktu nutná pro pasivní udržení znělosti.

Tyto závěry byly podloženy i měřením. Patricia Keating pro americkou angličtinu naměřila, že znělost u [b] trvá 55 ms, u [d] 40 ms a u [g] 31 ms (citováno v: Skarnitzl, Znělostní kontrast nejen v češtině, 2011, str. 56). Je zajímavé, že u českých exploziv nejsou měření tak jednoznačná. Machač ve své práci z roku 2006 uvádí, že průměrné hodnoty trvání znělosti jsou u [b] 70 ms, u [d] 47 ms a u [g] 60 ms (citováno v: Skarnitzl, Znělostní kontrast nejen v češtině, 2011, str. 58). Skarnitzl však dodává, že nízká hodnota u [d] může být způsobena jeho realizací. V češtině se [d] často v praxi realizuje jako alveolární švih [ɾ].

5. HLsyn

HLsyn je syntetizér, který bude v této práci k syntéze využíván. Leží na pomezí artikulační a formantové syntézy. Někdy je nazýván jako hybridní systém (Sensimetrics Corporation, 2004, str. 3) a někdy jako kvazi-artikulační syntetizér (Heid & Hawkins, 1998, str. 220). Definujícím prvkem tohoto syntetizéru jsou parametry, které uživatel pro syntézu zadává. Ty jsou složeny jak z parametrů vycházejících z fyziologického nastavení mluvidel, tak z parametrů, které vycházejí z požadavků na akustické parametry výsledného signálu. Je zřejmé, že se jedná o skloubení prvků artikulační a formantové syntézy.

5.1. Parametry

Syntetizér HLsyn pracuje celkem se třinácti parametry, ze kterých poslední tři byly přidány až od verze programu 2.2 (Sensimetrics Corporation, 2004, str. 2/supplement). Každý parametr je v grafickém uživatelském rozhraní zastoupen zkratkou. Pojďme si nyní všechny nastavitelné parametry v kostce popsat.

Area of glottis (*ag*) je parametr, který definuje plochu glottis v milimetrech čtverečních. Manipulováním tohoto parametru se můžeme pohybovat na škále od zcela zavřené glottis po plně otevřenou. V důsledku to bude znamenat možnost různých fonačních nastavení, jako například pro fonaci při hláskách znělých, otevřenou glottis pro hlásky neznělé, možnost glottis uzavřít pro vytvoření rázu a další.

Pomocí parametru Area of lip constriction (*al*) lze nastavit míru uzavření rtů, což je potřebné především při tvorbě hlásek labiálních. Údaj se opět zadává v milimetrech čtverečních. 0 mm^2 značí plné sevření a hodnota 100 mm^2 značí uvolněné otevřené rty s otvorem o ploše právě 100 mm^2.

Area of tongue blade constriction (*ab*) je údaj, pomocí kterého lze nastavit velikost plochy doteku jazyka s horním patrem. Tímto parametrem se manipuluje především při tvorbě koronálních hlásek.

Area of nasal opening (*an*) je parametr, který nastavuje oblast měkkého patra. V milimetrech čtverečních se zde nastavuje plocha otevření nazální dutiny. To se samozřejmě hodí ke tvorbě nazálních hlásek.

Parametr Rate of active change (*ue*) se udává v centimetrech krychlových a značí, o kolik se zvětší objem vokálního traktu před překážkou. Tohoto parametru se využívá při tvorbě obstruentů, kde dochází k plnému uzavření vokálního traktu. Aby bylo možno fonovat, tak se vokální trakt musí zvětšit, což je právě vyčísleno parametrem *ue*. Tento jev byl podrobněji popsán v oddílu 4.1.1.

Důležitý parametr je Fundamental frequency (*f0*), kterým lze nastavit frekvenci kmitání hlasivek. Hodnota se udává v desetinách hertzů bez desetinné čáry. Pro základní hlasivkovou frekvenci 150 Hz se tak vyplní údaj 1500.

Následují čtyři parametry odpovídající frekvencím prvních čtyř formantů (*f1, f2, f3, f4*; Frequency of 1st formant, 2nd formant, atd.). Ty se udávají opět v hertzích, tentokrát už ale v běžném celočíselném formátu. Pro první formant o výšce 500 Hz tak do pole *f1* vyplníme hodnotu 500.

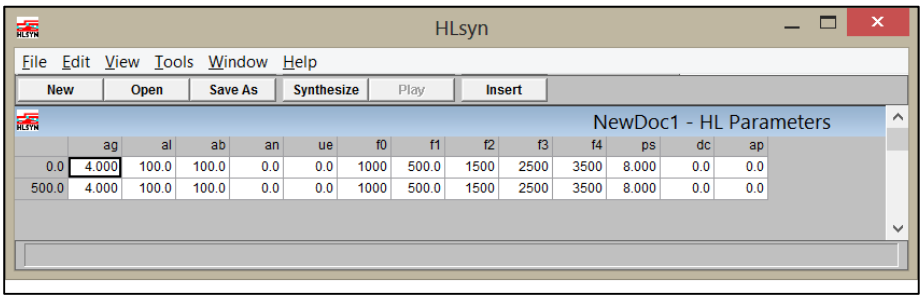
Následující tři parametry jsou platné pouze od verze HLsyn 2.2. Subglottal pressure (*ps*) byl přítomen i v předchozích verzích programu, avšak pouze jako konstanta neměnná v čase. Nově je možno tímto parametrem manipulovat v čase. Údaj specifikuje subglotální tlak a udává se jako výška vodního sloupce v centimetrech.

Delta compliance of the walls of the vocal tract (*dc*) je parametr, kterým lze nastavit poddajnost tkání ve vokálním traktu. Ve dřívější verzi HLsyn se zvlášť nastavovaly dvě hodnoty, které též vyústily v odpovídající hodnotu, ale opět pouze jako konstanta pro celou délku promluvy. Nově tedy vznikl parametr *dc*, který sdružuje obě hodnoty a dává uživateli možnost manipulovat s nimi v čase. Údaj se zadává v procentech, která značí odchylku od běžného stavu.

Hodnota Area of posterior glottal chink (*ap*) umožňuje uživateli lépe manipulovat se zdrojem šumu, pomocí kterého se dospěje k lepším výsledkům při syntéze znělých frikativ, nově umožní syntézu znělých aspirovaných exploziv a nakonec také pomůže vytvořit hlas, který je vnímán jako dyšný. Udává se v milimetrech čtverečných.

5.2. Práce v prostředí HLsyn

Uživatelské prostředí HLsyn se na první pohled tváří jako tabulkový procesor. Sloupce představují jednotlivé parametry a řádky představují časovou osu. V praxi to znamená, že pokud chceme syntetizovat například neutrální vokál [ə] o délce 500 milisekund, pak budeme operovat se dvěma řádky (Obrázek 7).



Obrázek 7 Základní pracovní okno programu HLsyn. Ukázka nastavení syntézy pro neutrální vokál [ə] o délce 500 milisekund.

První řádek bude odpovídat stavu v čase nula, druhý řádek pak stavu v čase 500 ms. Vyplníme všechny potřebné parametry pro oba řádky postupně (zde vycházíme ze základní obrazovky při spuštění programu, není tedy zapotřebí nic nastavovat). Úsek mezi bodem v čase 0 a 500 ms není třeba vyplňovat. Program automaticky vezme hodnoty mezi dvěma sousedícími řádky a hodnoty v mezičase doplní. Pokud jsou hodnoty ve sloupci stejné, pak použije stejnou hodnotu pro celou dobu trvání. Pokud se hodnoty liší, rozprostře potřebnou změnu rovnoměrně po celý časový úsek trvání (zde zmíněných 500 ms). Pokud jsme s nastavenými parametry spokojeni, stiskneme tlačítko „Synthesize“, čímž program spočítá námi zadané údaje a vygeneruje odpovídající zvuk. Následným stiskem tlačítka „Play“ je možno si ho poslechnout. V našem případě jsme vygenerovali 500 milisekund trvající neutrální vokál [ə].

Pokud se rozhodneme pro syntézu signálu, který se bude v čase měnit, což je i praxe každé přirozené lidské promluvy, je dobré si předem načrtnout základní časové milníky pro změnu parametrů pro syntézu. Program totiž sice umožňuje vkládat nové body změny pro libovolný čas, ale již vložené časové body nelze na časové ose posouvat. Jinými slovy, u každého řádku lze měnit veškeré parametry krom právě času. V praxi to často může znamenat, že změna jednoho časového údaje rovněž vede ke změnám všech následujících hodnot na časové ose. Není výjimkou, že se pak musí práce začít znovu od nuly. V případě nouze lze čas editovat také, to však bude popsáno níže. Nyní si pro ilustraci popíšeme dva příklady, kdy budeme manipulovat s parametry tak, abychom syntetizovali signál měnící se v čase.

První příklad bude ilustrovat jednoduchou změnu f0, která se bude v čase postupně zvyšovat z hodnoty 120 Hz na hodnotu 225 Hz. Vyjdeme ze základního nastavení programu po spuštění. V čase 0 ms vložíme hodnotu do sloupce označeného *f0* 1200 a v čase 500 ms hodnotu 2250. Stiskneme tlačítko „Synthesize“ a následně „Play“. Nyní můžeme slyšet postupné zvyšování základní frekvence.

V druhém příkladu budeme chtít stejnou změnu, ovšem zvýšení hodnoty f0 budeme chtít realizovat skokově v čase 250 ms. Vyjdeme z předchozího nastavení a pomocí tlačítka „Insert“ vložíme řádek v čase 250 ms. Hodnotu *f0* vyplníme na požadovaných 225 Hz. Abychom dosáhli stanoveného cíle, musíme ještě vložit řádek s časem 249 ms a v tom nastavit hodnotu f0 na 120 Hz. Pokud bychom tak neučinili, prvních 250 ms by hodnota f0 postupně stoupala na 225 Hz a druhých 250 ms by na této hodnotě zůstala. Pokud však „ukotvíme“ hodnotu 120 Hz v čase 249 ms vloženým řádkem, bude prvních 250 ms generováno s hodnotou f0 120 Hz a druhých 250 ms s hodnotou 225 Hz, jak jsme si zadali. Po stisknutí tlačítka „Synthetize“ a následně „Play“ si můžeme poslechnout výsledek.

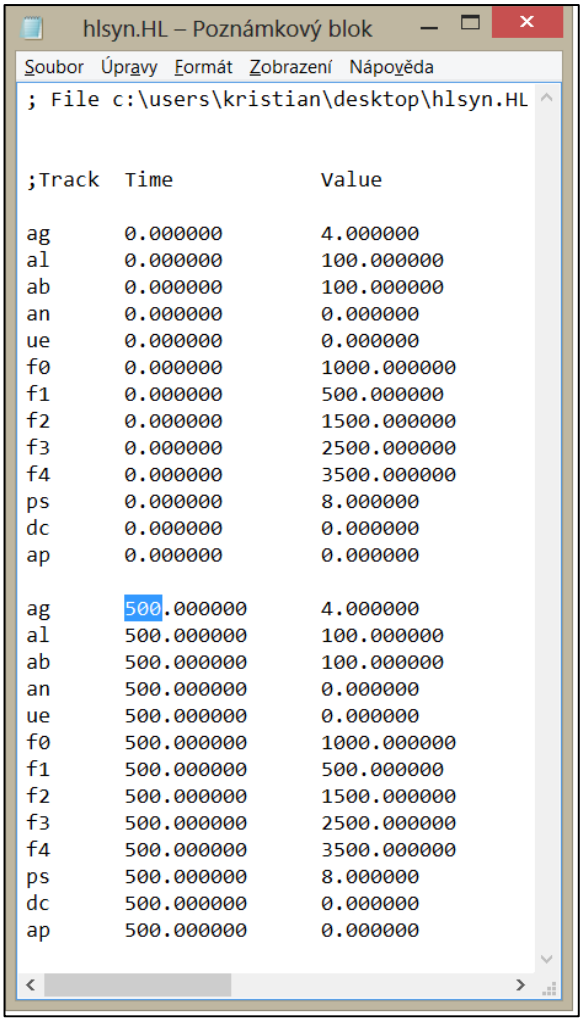
Po jednoduché ilustraci práce v programu je zřejmé, že je po celou dobu nutné mít poměrně jasnou představu o tom, co chceme syntetizovat. Konkrétně jde o představu toho,

jaké změny v jakém čase probíhají při produkci požadovaných hlásek. Nejedná se tedy o jednoduchou syntézu, kde by se zadávaly rovnou hlásky, a program by sám doplňoval údaje. Zároveň je nutné myslet na přechody mezi hláskami, které se svou přítomností v okolí často ovlivňují. Tyto koartikulační změny je nutné též brát v potaz při editaci parametrů, neboť HLsyn je nijak nepředjímá. Nevýhodou tohoto pracovního postupu je jeho zdlouhavost a mnoho prostoru pro chybné vyplnění údajů a tedy snížení přirozenosti. Na druhou stranu je kladem, že lze velmi přesně nasimulovat přesně takovou situaci, jakou chceme. Můžeme manipulovat právě těmi charakteristikami v signálu, na které cílíme. Například můžeme manipulovat právě se znělostní charakteristikou a sledovat, jaké změny to pro posluchače ponese. Ostatní parametry zůstávají stejné a my tak máme absolutní kontrolu nad tím, jak výsledný signál vypadá. Podobná manipulace se signálem živého mluvčího by byla přinejmenším problematická.

5.3. Editování času

Jak bylo řečeno již dříve, program HLsyn neumožňuje editovat čas u jednotlivých řádků. Při syntéze se však často hledá nejlepší nastavení parametrů tak, aby signál zněl co nevíce přirozeně. Často je větší představa o tom, jaké nastavení parametrů použít, než v jakém čase ho použít. Naštěstí existuje způsob, jak programu vnutit změnu času. Není to standardní postup popsaný v manuálu, proto je vhodné ho zde blíže popsat.

Nejprve je nutné hodnoty exportovat do



Obrázek 8 Ukázka editace parametrů HLsyn v Poznámkovém bloku.

tzv. human-readable formátu. V hlavním menu zvolíme položku File-Export-HL Parameters. Zvolíme místo uložení dle naší preference. Pokud bychom v tomto bodě uložený soubor otevřeli v programu Microsoft Excel 2013, tak by se špatně detekovaly sloupce a hodnoty času by bohužel nebyly v samostatném sloupci. Po editaci a uložení by navíc při importu zahlásil HLsyn chybu. Je nutné použít program Poznámkový blok nebo jeho preferovanou alternativu a otevřít uložený soubor v něm. Pak máme na výběr dvě možnosti. Pokud chceme jednorázově přepsat jeden řádek, pak stačí požadovaný čas ručně přepsat rovnou v Poznámkovém boku (viz Obrázek 8). Soubor uložíme a v HLsyn zvolíme File-Import-HL Parametres a zvolíme námi editovaný soubor.

V případě, že budeme chtít s časem manipulovat vícekrát, trochu práce si lze ušetřit zkopírováním celého textu z Poznámkového bloku do Excelu s použitím Průvodce vkládání textu. Jako oddělovač zvolíme mezery a tabelátory. Pak můžeme použít šikovné funkce Najít a nahradit. Po nahrazení časových údajů novými zkopírujeme veškerý obsah zpět do Poznámkového bloku a v tom soubor uložíme. Po importu HLsyn načte hodnoty bez chybové hlášky.

Podrobnější popis páce s importem textových hodnot do MS Excel je popsán v nápovědě příslušného programu. Tento postup je vyzkoušen pro verzi Windows 8, MS Office 2013 a HLsyn 2.2. Jeho funkčnost v jiných verzích programů se může lišit.

6. Vlastní syntéza

Pro potřeby této práce byly použity nosné fráze, do kterých byla vsazována sledovaná kvazi-slova. Celá nosná fráze byla z důvodu co nejmenšího rušení posluchače syntetizována rovněž pomocí HLsyn stejně tak jako kvazi-slova. Nosná fráze byla zvolena s ohledem na to, aby se sledovaná kvazi-slova vyskytovala v mediální pozici a celá fráze přitom dávala smysl. Nakonec byla zvolena fráze „Zopakuj aXa prosím“, kde „aXa“ představuje sledované kvazi-slovo, které má strukturu vokál-konsonant-vokál. Konsonant byla vždy některá z českých explozív a vokál byl reprezentován českým vokálem [a]. U exploziv bylo v programu HLsyn manipulováno se znělostí, čímž vzniklo několik variant daného kvazi-slova. Při vytváření vokálů se vycházelo ze dvou zdrojů. Pro syntézu vokálů byla výchozím bodem studie Skarnitzla a Volína (Skarnitzl & Volín, 2012). Pro většinu ostatních hodnot byl odrazovým můstkem studijní materiál kurzu na procvičení práce s programem HLsyn (Sensimetrics Corporation, 1995). S přihlédnutím k tomu, že Skarnitzl a Volín zkoumali vokálové formanty spíše z běžné promluvy neprofesionálních mluvčích, kde nebyl kladen důraz na precizní výslovnost, bylo nutné hodnoty formantů ještě upravit tak, aby reprezentovali ortoepickou výslovnost českých vokálů. Drobné úpravy byly nutné i u exploziv, jelikož studijní materiál k programu HLsyn je zaměřen spíše na anglické fonémy. Konkrétní úpravy budou rozebrány později v této práci. Syntetizovaná kvazi-slova spolu s nosnou frází byla následně poskládána do jednoho zvukového záznamu pomocí programu Cool Edit (Syntrillium Software Corporation, 2003). Takto pořízený záznam byl přehráván posluchačům, kteří měli za úkol do předem rozdaných dotazníků vyplnit, co slyší. Ukázka dotazníku je v příloze (Příloha A).

6.1. Nosná fráze

Požadavky na nosnou frázi z hlediska percepce byly především srozumitelnost a přirozenost. Pokud by byla fráze syntetizována striktně dle doporučovaných parametrů v manuálu HLsyn, zněla by značně strojeně. Jednak se u nosné fráze muselo dbát na

přirozenou intonaci a dále pak přihlédnout k reálné délce trvání jednotlivých fonémů. Z těchto důvodů se nejprve pořídila zkušební nahrávka dané fráze od živého mluvčího, která posloužila jako odrazový můstek pro délku trvání jednotlivých hlásek a rovněž jako vzor kontury f0, díky které mohlo být dosaženo přirozenější intonace. Intonace pro slovo „zopakuj“ byla zvolena jako neukončená a pro slovo „prosím“ byla zvolena intonace klesavá. Při vložení sledovaného kvazi-slova do nosné fráze pak vzniknul dojem jednolité fráze s přirozenou intonací. Mezery mezi jednotlivými slovy byly po pečlivém poslechu stanoveny arbitrárně na 65 milisekund po „zopakuj“ a na 95 milisekund před „prosím“.

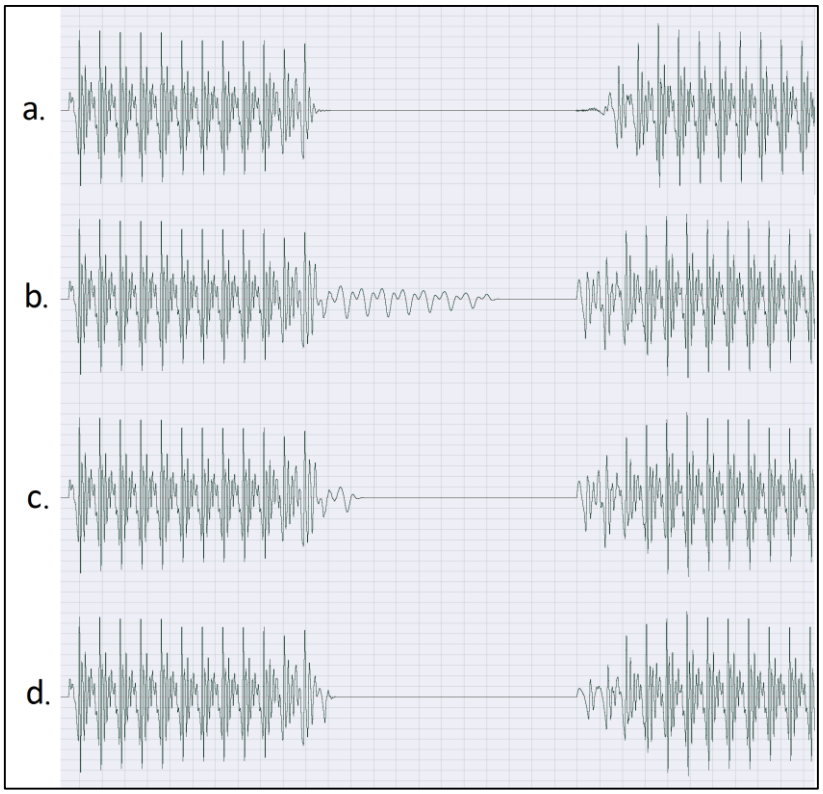
6.2. Kvazi-slova

Základní struktura kvazi-slov byla zmíněna již výše – aXa. Arbitrárně zvolený český vokál [a] předchází a následuje sledovanou českou explozivu, která se tak nachází v ideální intervokalické pozici. Požadavky na kvazi-slova byly však odlišné od nosné fráze. Zde byla vyžadována co nejpreciznější výslovnost a naopak plochá intonační křivka. Proto byla kvazi-slova syntetizována přímo v programu HLsyn bez jakýchkoliv orientačních předloh od živého mluvčího. Zadávání hodnot pro syntézu probíhalo ve dvou hlavních fázích. Nejprve byly zadány předpisové hodnoty doporučované tvůrci programu HLsyn. Protože syntetizované explozivy poslechově nedosahovaly potřebné kvality, bylo nutné provést druhou vlnu úprav. V té se manipulovalo především s délkou intenzní a detenzní části exploziv s ohledem na místo artikulace. V některých případech bylo nutné hýbat i s hodnotami jednotlivých formantů. Jako zvlášť problematická se ukázala syntéza exploziv [d] a [t]. Především u znělé varianty poslechově docházelo ke znatelnému odchýlení od místa artikulace a to buď směrem k dentalizaci nebo k platalizaci. Hodnoty navrhované v manuálu HLsyn se pro české [d] a [t] ukázaly jako značně nevyhovující. Bohužel se nepodařilo nalézt takové hodnoty pro syntézu, aby byla v poslechu deformace odstraněna, ale alespoň došlo k jejímu výraznému utlumení. V případě neznělé varianty byl problém obdobný, ale poslechově ne již tak výrazný.

6.3. Manipulace se znělostí

S ohledem na parametry zadávané v programu HLsyn lze se znělostí manipulovat v několika směrech. Jednak lze hýbat s parametrem *ue*, který má vliv na zvětšování nadhrtanových částí vokálního traktu tak, aby byl kompenzován závěr, který by jinak bránil kmitání hlasivek. Dále lze manipulovat s parametrem *ag*, který se přímo podílí na generování hlasivkového tónu. Poslední teoretickou možností je znatelně zkrátit trvání závěrové části explozivy. Je určité podezření, že zkrácení může u posluchače subjektivně vést ke ztrátě znělosti.

V rámci této práce bylo použito všech výše zmíněných možností způsobem, který nyní bude popsán blíže. Každý znělostní pár vždy obsahoval zcela neznělou a zcela znělou variantu explozivy tak, jak se přirozeně vyskytují v českém jazyce. Dále pak byla syntetizována taková exploziva, kde nebyl použit parametr *ue*. Jinými slovy byl parametr po celou dobu trvání ponechán na hodnotě *ue* = 0, čímž se podstatným způsobem zkrátila znělost v závěrové části explozivy. Zmiňme zde pouze odkaz na oddíl 4.1.1, kde byl tento mechanismus podrobně demonstrován. Čtvrtou variantou byla exploziva, u které bylo ponecháno nastavení parametru *ue* odpovídající zvětšení nadhrtanových dutin, ale byl změněn parametr *ag*. Byla použita stejná hodnota,



Obrázek 9 Oscilogram kvazi-slov s bilabiální explozivou **a.** znělou, **b.** neznělou, **c.** znělou (artikulace bez zvětšení nadhrtanových dutin), **d.** znělou (nižší podíl tónové složky).

jaká se v programu HLsyn používá pro znělé frikativy, čímž došlo k výraznému snížení tónové složky v signálu. Namísto hodnoty *ag* = 4, která odpovídá většině znělých hlásek, byla použita hodnota *ag* = 10. Tato hodnota se doporučuje používat u znělých frikativ, neboť lépe odráží jejich specifickou artikulaci. Stejná hodnota byla použita i v hlásce [z] ve slově „zopakuj“ v nosné frázi. V páté variantě explozivy bylo manipulováno s parametrem času tak, že došlo ke zkrácení intenzní části explozivy o 30 milisekund. Všechny ostatní parametry zůstaly nezměněny. Poslední, šestá, varianta byla opět o 30 milisekund zkrácená exploziva, navíc však s vynechaným nastavením parametru *ue*, jak tomu bylo učiněno u varianty číslo tři. Přehledný pohled na prováděné změny poskytuje tabulka v příloze (Příloha B), reálné manifestace změn znělosti v signálu lze pozorovat na obrázku (Obrázek 9).

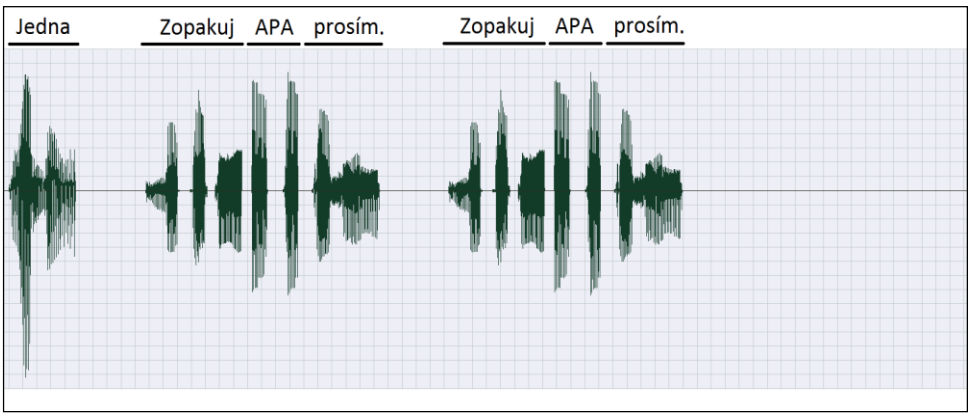
6.4. Percepční test

Vlastní percepční test trval necelých pět minut a obsahoval dohromady 36 položek. To znamená 6 variant ke každému ze tří sledovaných míst artikulace ve dvou opakováních. Pořadí položek v testu bylo zvoleno kvazi náhodně. Nejprve bylo pomocí funkce náhodné číslo v programu MS Excel sestaveno předběžné pořadí položek do testu, které bylo následně ručně upraveno tak, aby se za sebou nikdy nevyskytovaly dvě explozivy se stejným místem tvoření. Tím vzniklo finální pořadí položek v testu.

Vlastní položka v poslechovém testu obsahovala nosnou frázi s vsazeným sledovaným kvazi-slovem. Celá tato fráze byla dvakrát za sebou zopakována. Dále byla každá položka uvozena svým pořadovým číslem, které bylo namluveno školeným lidským mluvčím. Na konci každé položky byl ponechán krátký čas na odpověď respondenta. V rámci všech syntetizovaných položek bylo sledováno, zdali se při přehrání netvoří nechtěné artefakty na konci signálu. Pokud tomu tak bylo, vždy byla umazána poslední vlna v signálu tak, aby byl tento artefakt odstraněn. Vždy šlo o nejmenší možnou část signálu, kdy délka vymazané části

často nedosahovala ani 1 milisekundy. Názorný pohled na průběh položky v percepčním testu viz obrázek (Obrázek 10).

Respondenti své odpovědi zapisovali do předem vytištěného odpovědního archu, ve kterém měli pod číslem položky vždy na výběr tři možnosti: konkrétní znělou variantu explozivy, konkrétní neznělou variantu explozivy a možnost „něco mezi“, kterou měli respondenti označit v případě, že slyší hlásku na pomezí znělé a neznělé.



Obrázek 10 Zobrazení průběhu konkrétní položky v poslechovém testu.

Každý respondent obdržel instrukce k poslechu a odpovědní arch. Instrukce jasně popisovaly zadání úkolu a každý respondent měl možnost se v případě nepochopení před testem dotázat. Pro dobré pochopení byly na začátek percepčního testu zařazeny i tři zácvičné položky, které respondent viděl v odpovědním archu vyplněné. Měl tak možnost slyšet variantu znělou, neznělou a částečně znělou a zároveň si v případě potřeby ještě upravit hlasitost poslechu. Krom průběhu samotného poslechového testu byl každý respondent poučen i o tom, že poslech smí vykonat pouze v klidném prostředí s nasazenými circumaurálními sluchátky. Dále mu bylo vysvětleno, že mu bude reprodukována syntetizovaná řeč, tudíž mu může připadat více či méně přirozená v porovnání s živým mluvčím, na což ale neměl brát posluchač nijak zřetel. Poslední důležitý bod instrukcí zdůrazňoval to, že percepční test je vystavěn tak, že není předem dáno, co je správná a co špatná odpověď a tudíž se respondent nemusí bát zaškrtnout to, co skutečně slyší. To mělo za úkol sejmout z respondenta případný stres z hodnocení jeho poslechových nebo jiných schopností.

Součástí odpovědního archu byly i identifikační údaje respondenta, na základě kterých mu byl přidělen unikátní kód pro pozdější práci s daty. Respondent měl za úkol vyplnit své iniciály, pohlaví a rok narození. Pokud vezmeme jako příklad hypotetického respondenta se jménem Kristian Urban, který se narodil roku 1984, pak na základě jím poskytnutých informací mu byl přidělen kód KU84m.

7. Výsledky

U všech respondentů byly sledovány dva faktory, pomocí kterých byla vyhodnocena důvěryhodnost každého z nich. Prvním faktorem byla míra shody v odpovědích. Každé unikátní kvazi-slovo v testu bylo zopakováno v rámci dvou položek. Odpovědi v těchto navzájem identických položkách byly u každého respondenta porovnány a byla vyhodnocena míra jejich shody neboli konzistence respondenta. Tento faktor měl za úkol odhalit nepozorné posluchače a ty, kteří by test vyplňovali, aniž by si test skutečně poslechli.

Druhý faktor je faktor shody znělých a neznělých. Každý respondent slyšel celkově šest kvazi-slov se zcela znělou a šest se zcela neznělou explozivou. Tento faktor měl za úkol odhalit posluchače, kteří by kvalitou svého sluchu výrazně vybočovali z běžného průměru ostatních respondentů. Zároveň bylo přistupováno jiným způsobem k položkám znělým a k položkám neznělým. Větší pozornost byla věnována těm respondentům, kteří označovali neznělé explozivy jako znělé než naopak. Zatímco u znělých může být posluchač v průběhu testu zmaten, zdali se náhodou nejedná o částečnou znělost, u neznělých by respondent v signálu slyšel něco, co tam objektivně není. V praxi se mimo jiné skutečně ukázalo, že neznělé explozivy byly daleko odolnější k chybnému hodnocení než ty znělé.

Pokud byl některý z respondentů zachycen výše pospaným mechanismem ať už v rámci prvního nebo druhého faktoru, byla mu věnována větší pozornost. Především byl respondent kontaktován a dotázán na průběh jeho poslechu a jeho subjektivní pocit z testu. Na základě výše zmíněného kontrolního postupu bylo nutné z výsledků vyřadit tři respondenty pod kódovým označením MŠ46f, JB86m a ML82m. U prvního respondenta byla dodatečně zjištěna sluchová vada a užívání naslouchátka. Tomu napovídaly i výsledky, které vykazovaly nejnižší konzistenci ze všech respondentů. Respondent rovněž vykazoval nejnižší míru shody se stoprocentně neznělými explozivami, kdy pouze jedna neznělá byla označena jako skutečně neznělá. Zde by mohl být proveden zajímavý výzkum toho, jestli izolovaně v intervokalické pozici sice respondent slyší explozivy jako znělé, ale v přirozené promluvě

dle významu doplňuje znělost a neznělost až mozek jako korekční funkci. To je však teorie nad rámec této práce a nebudeme se jí zde dále věnovat.

U druhého respondenta žádná sluchová vada zjištěna nebyla, ale při delším rozhovoru bylo zjištěno subjektivní podezření na blíže nespecifikovanou řečovou vadu. Krom toho byl v historii respondenta odhalen nedávný úraz v oblasti hlavy. Jeho výsledky sice vykazovaly nadprůměrnou konzistenci, ta ale byla způsobena tím, že prakticky vše označoval jako znělé. Sám respondent si toto uvědomoval a potvrdil, že slyšel ve většině případů znělou variantu. Jeho výsledky i tak ale do výsledků zařazeny nebyly.

Třetí respondent se po dotazu k průběhu testu dodatečně přiznal, že v době poslechu trpěl rýmou a zalehlýma ušima. I jeho výsledky vykazovaly znatelnou odchylku od výsledků ostatních respondentů. Zbylí respondenti již požadavky splnili a byli zahrnuti do výsledků. Celkem byly zahrnuty odpovědi od 18 respondentů náhodného složení ve věkové kategorii od 26 do 64 let. Všichni respondenti zahrnutí do výsledků potvrdili, že si nejsou vědomi žádné sluchové vady, kterou by v době poslechového testu trpěli.

Pro lepší srovnání znělostních charakteristik se v této práci bude pracovat s mechanismem, který zde budeme nazývat znělostní index. Je to arbitrárně zvolený ukazatel, který není primárně určen pro širší využití mimo tuto práci. Každá položka, kterou respondent ohodnotil jako znělou, byla v seznamu výsledků ohodnocena jedním bodem. Každá položka, která byla hodnocena jako neznělá, dostala přiřazenu hodnotu 0 a položka, která byla hodnocena jako částečně znělá, dostala připsanou hodnotu 0,5. Pokud se udělá pro sledovanou skupinu hlásek aritmetický průměr přidělených hodnot, dostaneme číslo v rozmezí od 0 do 1. Čím je číslo vyšší, tím více byla hláska hodnocena jako znělá a naopak. Nutno zde však podotknout, že tímto indexem nelze hodnotit skupinu respondentů jako takovou s úmyslem zjistit, zda mají obecnou tendenci slyšet spíše znělé či spíše neznělé hlásky. Není totiž možné tvrdit, že v testu bylo zařazené stejné množství neznělých jako znělých hlásek a tudíž by měl být celkový index všech respondentů u všech hlásek ideálně

0,5. Plně znělých a neznělých hlásek skutečně stejný počet byl, ale u manipulovaných hlásek nelze stanovit, že ta či ona položka by měla být vnímána jako znělá nebo jako neznělá.

Když se blíže podíváme na explozivy, u kterých byla syntetizována plná neznělost a plná znělost, zjistíme, že výsledky naplňují očekávání. První jmenovaná skupina dosáhla indexu znělosti 0,03. V číslech to znamená, že z celkového výskytu 108 hodnocených položek bez znělosti byly pouze v 6 výskytech hodnoceny respondenty jinak než neznělé. Ve všech těchto čtyřech případech byly hodnoceny jako částečně znělé. Podobně očekávání naplnily i hlásky zcela znělé, které dosáhly indexu znělosti 0,94. Jinak než znělé byly hodnoceny pouze v 10 z celkem 108 výskytů. Z těchto 10 výskytů pak bylo hodnoceno 8 položek jako částečně znělých a 2 výskyty hodnotil jeden respondent jako neznělé.

Zajímavější je samozřejmě pohled na hlásky, kde bylo se znělostí manipulováno. Explozivy, u kterých bylo manipulováno s parametrem *ue*, se umístily v pomyslném středu vnímané znělosti, neboť dosáhly indexu znělosti 0,56. Z celkového počtu 108 výskytů hlásek s vypuštěným parametrem *ue* bylo 52 hodnoceno jako částečně znělých, 34 jako plně znělých a 21 jako neznělých. To znamená, že v 48% případů byla exploziva hodnocena jako částečně znělá, v 32% případů jako plně znělá a v 20% případů jako neznělá.

Manipulace s parametrem *ag* přinesla ve srovnání s výsledky u manipulace s parametrem *ue* výraznější trend jedním směrem. Index znělosti je v tomto případě 0,18 a značí tak, že znělost je v tomto případě vnímána téměř výjimečně. 77 položek ze 108 bylo hodnoceno jako neznělé a pouze 7 byl hodnoceno jako znělé. Zbylých 24 položek bylo hodnoceno jako částečně znělé. Lze tak zkonstatovat, že mírné oddálení hlasivek při fonaci (a tedy snížení podílu tónové složky) má výrazně vyšší vliv na vnímané oslabení znělosti než statická poloha hrtanu. To by poukazovalo na to, že pohyb hrtanu má za následek umožnit přesnější identifikaci příznaku znělosti. Bez tohoto mechanismu by muselo být vnímání znělosti citlivější, a tedy náročnější, nebo by se příznak znělosti musel rozlišovat čistě dle kontextu.

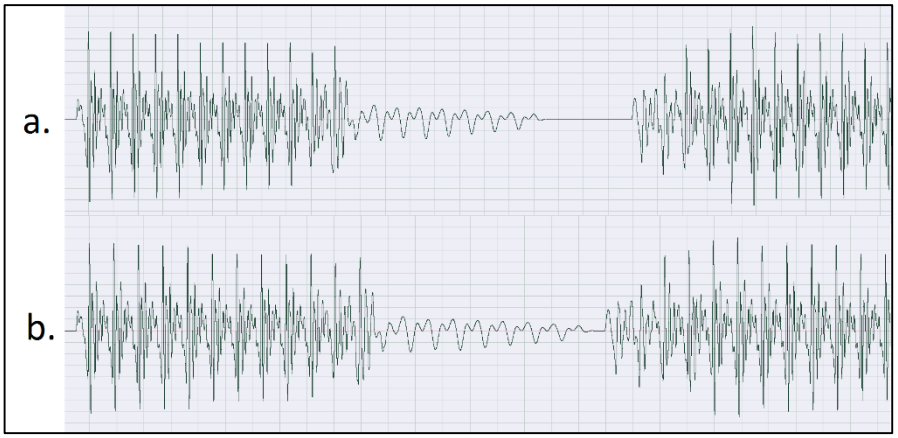
Druhý zmíněný případ by však svým způsobem dával zaniknout bipolárnímu rozdělení na explozivy znělé a neznělé a opozice znělosti by zanikla.

Přitom tak značná míra dopadu zvýšení hodnoty parametru *ag* na snížení percepční znělosti nemohla být zcela očekávatelná. Při manipulaci byl použit parametr *ag* = 10 mm^2, přičemž základní hodnota pro znělé hlásky je *ag* = 4 mm^2. Zajímavé je srovnání se znělými frikativami, kde byla v nosné frázi použita hodnota parametru *ag* = 10 mm^2, která se při percepci subjektivně zdála jako optimální. Tvůrci syntetizéru HLsyn však navrhují pro znělé frikativy hodnotu až *ag* = 20 mm^2, což znamená další výrazné oslabení znělosti. Z toho vyplývá, že frikativám k percepčnímu udržení znělosti stačí daleko méně tónové složky než explozivám. Lze se domnívat, že zatímco s námi testovanou hodnotou parametru *ag* ještě někteří respondenti znělost vnímali, s hodnotou ag = 20 mm^2 by byly explozívy vnímány pouze jako neznělé.

Při pohledu na manipulaci s parametrem času se ukázalo, že v obou případech se znělostní charakteristiky hlásek pouze posilují. Varianta, ve které byla pouze zkrácena tenzní část explozivy, vykazuje index znělosti 0,97, což v číslech představuje, že oproti původním 10 položkám bylo již jen 5 hodnoceno jako ne plně znělých. Varianta, kde byl zkrácen čas tenzní části a zároveň nebyl použit parametr *ue* pro kompenzaci znělosti, získala index znělosti 0,70. Respondenty byla hodnocena v 50 případech jako znělá, v 52 případech jako částečně znělá a jen v 6 případech jako neznělá. Zatímco částečná znělost byla respondenty zvolena v porovnání s nezkrácenou variantou téhož v přibližně stejném počtu případů, znatelně se změnil poměr plně znělých k plně neznělým ve prospěch znělých.

Zkrácení hlásky za podmínek daných touto prací k oslabení vnímané znělosti nevedlo, tendence byla ve skutečnosti opačná. Možné vysvětlení se nabízí v tom, že pokud je zkrácen závěr, tak se též prodlouží doba, po kterou mohou hlasivky při tvorbě znělé explozivy kmitat. Nestihne se totiž plně projevit efekt vyrovnání tlaků pod hlasivkami a nad hlasivkami. Tím je umožněno hlasivkám kmitat po delší poměrnou část závěru než u explozivy s delší tenzní

částí. Lze očekávat, že tento efekt bude percepčně výraznější v případě, že je nadhrtanovým dutinám umožněno se dočasně zvětšit než opačně. Tím by se i vysvětlilo to, že posílení znělosti při statickém objemu nadhrtanových dutin je výraznější než pokud se objem nadhrtanových dutin zvětší. Pohled na průběh znělosti u znělé bilabiální explozivy a její varianty se zkrácenou tenzní částí lze vidět na obrázku (Obrázek 11).



Obrázek 11 Oscilogram bilabiální explozivy **a.** znělé, **b.** znělé se zkrácenou tenzní částí. Na oscilogramu lze pozorovat, že v druhém případě je znělost téměř po celou dobu tenze. Ve skutečnosti je v obou případech znělost v tenzní části stejně dlouhá, jen v druhém případě je kratší samotná tenze (časová osa je u obou ukázek mírně odlišná, ukázka **b.** je o 30 ms kratší).

Celkově se pak ukázalo, že explozivy se dle znělosti rozmístily do třech od sebe odlišitelných skupin. Indexy 0,97 a 0,94 tvořící percepčně znělé explozivy, indexy 0,70 a 0,56 tvořící percepčně částečně znělé explozivy a nakonec indexy 0,18 a 0,03 tvoří skupinu percepčně neznělých exploziv.

Zajímavé je i srovnání toho, jak stabilní je vnímaná znělost v jednotlivých místech tvoření. Obecný předpoklad je, že čím je místo tvoření dále od hrtanu, tím je více prostoru pro zachování znělosti. V našem případě se toto však ne zcela potvrdilo. Nejvyšší index znělosti, konkrétně 0,60, vykazuje bilabiální artikulace. Toto umístění je ještě dle očekávání a lze konstatovat, že na základě našich měření mají bilabiální explozivy schopnost zachovat si znělost vyšší než zbylé dvě sledované skupiny. Druhá nejstabilnější pozice však dle našich měření připadla velárním explozivám s indexem znělosti 0,57. Alveolární explozivy dosáhly indexu 0,53.

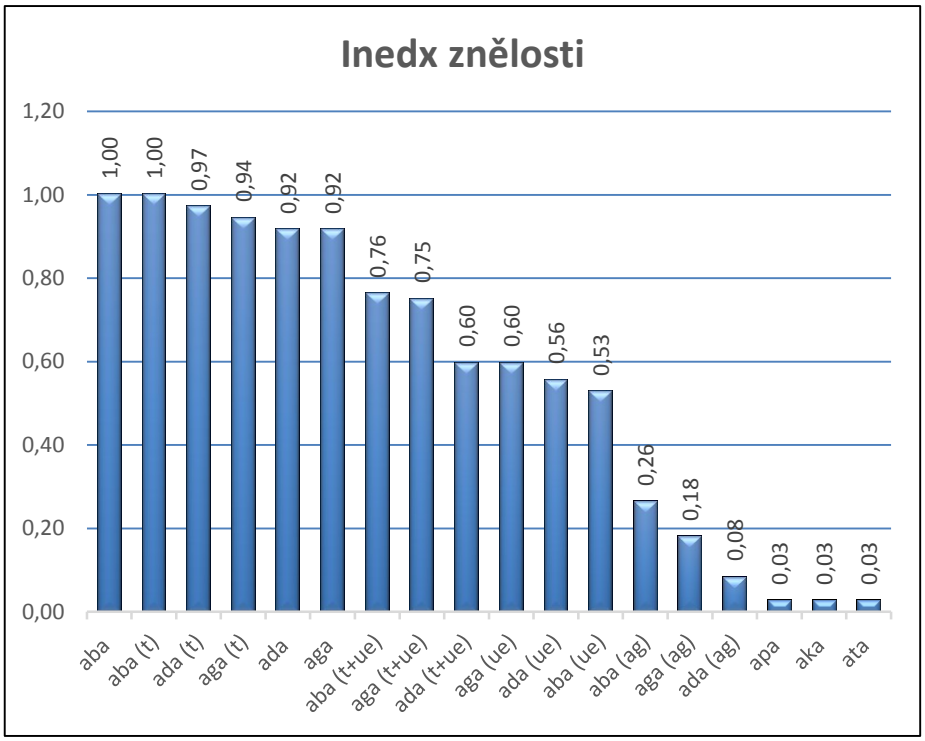
Proč druhé dvě pozice nenaplňují očekávání, by mohlo jít přisoudit tomu, že za pomoci syntetizéru HLsyn nebylo možno syntetizovat alveolární explozivy, aniž by v nich percepčně

nezůstal určitý náznak dentalizace. České hlásky [t] a [d] se liší nejen zněloství, ale částečně i místem artikulace. Právě neznělá varianta je obvykle artikulována blíže na zubech, což mohlo v posluchačích vyvolat často pochyby u znělé varianty a přiklonit je více k variantě neznělé. Ačkoliv tato rozdílnost byla v nastavení reflektována a byla snaha obě hlásky co nejvíce přiblížit ortoepické výslovnosti, je nutno připustit, že se tak stalo jen do určité míry. Ačkoliv byli respondenti upozorněni, ať neberou na případnou nepřirozenost syntézy zřetel, může být toto faktor, který výsledky ovlivnil.

V návaznosti na předchozí odstavec se lze podívat i na znělostní index jednotlivých míst tvoření pouze ve skupině hlásek bez manipulované znělosti a pro porovnání pak i hlásek s manipulovanou znělostí. V případě hlásek bez manipulace vychází nejvyšší index znělosti opět u bilabiálních exploziv (0,51), avšak u velárních a alveolárních se již rozdíl smazal a obě skupiny se umístily se shodným indexem 0,47. Tyto hodnoty se už více přibližují tomu, co lze teoreticky očekávat. U hlásek manipulovaných zůstává pořadí nezměněno. Nejvíce znělé se respondentům jevily bilabiály, v závěsu pak veláry a s vyšším odstupem pak alveoláry. Řečí čísel to představuje indexy 0,64, 0,62 a 0,55. S ohledem na výsledky v této práci lze tedy pouze konstatovat, že bilabiální explozivy mají schopnost si lépe zachovat parametr znělosti oproti explozivám alveolárním a velárním. Co se posledních dvou jmenovaných týče, buď se projevila odlišnost českého [t] a [d] nebo zde lze sledovat náznak toho, že český posluchač je na rozdíly znělosti u velárních exploziv více citlivý, neboť znělost alveolárních exploziv je podepřena ještě specifickou artikulací.

Indexem znělosti lze porovnat i jednotlivé varianty exploziv proti sobě. Indexem 1,0 byly ohodnoceny položky se znělou bilabiální explozivou jak ve své nezměněné podobě tak ve variantě se zkrácenou tenzní částí. Žádná jiná exploziva nebyla ve všech položkách hodnocena jako znělá. Naproti tomu ani jediná položka nebyla hodnocena jako vždy neznělá. Je však zajímavé, že všechny neznělé varianty exploziv byly shodně hodnoceny stejným

indexem 0,03. U ostatních exploziv žádné nezvyklé tendence zjištěny nebyly a odrážejí výsledky, které byly vypsány v předchozích odstavcích. Podrobnější pohled nabízí tabulka (Tabulka 2).



Tabulka 2 Index znělosti pro jednotlivé explozivy. V závorce je uváděn typ manipulace, která byla s explozivou prováděna.

8. Závěr a diskuse

Dle očekávání byly neznělé explozivy vnímány jako neznělé a znělé explozivy jako znělé. Explozivy, ve kterých bylo manipulováno se znělostí, si percepčně udržovaly znělost různou měrou. Nejvíce znělost ztrácely ty explozivy, u kterých byla snížena tónová složka pomocí parametru *ag*. Ukázalo se, že explozivy mají při snížení tónové složky daleko vyšší pravděpodobnost ztráty vnímané znělosti ve srovnání s frikativami.

Manipulace s parametrem *ue*, ukazuje, že zvětšování objemu nadhrtanových dutin při produkci znělých exploziv má své reálné opodstatnění. Naše předběžné výsledky ukazují, že posluchač při absenci změny objemu nadhrtanových dutin pochybuje o tom, zda slyší znělou či neznělou explozivu. Jen některým taková míra znělosti stačila k označení explozivy jako znělé. Naproti tomu se ale začali vyskytovat v nezanedbatelné míře posluchači, kteří explozivu hodnotili jako neznělou, což byl v případě zachování mechanismu zvětšování objemu nadhrtanových dutin jev ojedinělý.

Manipulace s délkou trvání tenzní části znělých exploziv neprokázala žádný vliv na ztrátu znělosti. Naopak posilování znělosti při zkracování explozivy poukázalo na následující: pokud by se percepční oslabení znělosti exploziv v závislosti na jejich zkracování v rychlé promluvě potvrdilo, znamenalo by to, že nastavení mluvidel, výdechového proudu vzduchu a činnosti hlasivek jsou zásadně odlišné od běžného tempa produkce řeči. To by se pro další zkoumání tohoto jevu muselo odrazit v jiném nastavení parametrů syntézy. Lze předpokládat, že nejvíce by se to dotklo parametru *ue*, který by mohl být ovlivněn jak v délce svého trvání, tak i ve své intenzitě. Otázkou zůstává, zda by oslabená percepční znělost nebyla spojena také s obecně oslabenou artikulací, při které nejsou dosahována ideální místa artikulace, a zda a jak toto reflektovat v nastavení parametrů syntetizéru.

Stabilita udržení znělosti pro explozivy s místem tvoření závěru dále od hlasivek se sice v této práci potvrdila jen částečně, ale předpokládaný trend je přesto zřetelný. Horší rozlišení velárních a alveolárních exploziv bylo pravděpodobně způsobeno použitým syntetizérem

HLsyn, jehož ideální nastavení se pro české alveoláry nepodařilo nalézt. Pokud by měl být syntetizér HLsyn použit i pro další percepční ověřování českých hlásek, bylo by jistě na místě zpracovat samostatný percepční test zaměřený na nastavení parametrů českých alveolárních exploziv. V případě, že by se podařilo nalézt nastavení pro percepčně věrnou artikulaci alveolárních exploziv, bylo by na místě naše percepční ověření znělostních charakteristik exploziv zopakovat a udělat relevantnější závěry. Pokud by se takové nastavení nalézt nepodařilo, bylo by na místě zvážit další použití tohoto syntetizéru pro jakýkoliv výzkum, který by se dotčených hlásek týkal.

Vzhledem k předloženým výsledkům není překvapivé, že nejvíce si znělost udržela bilabiální explozíva [b]. Co je ale zajímavé, že to byla jediná exploziva, u které respondenti nikdy nepřiřadili jinou hodnotu než „znělá“. Zatímco neznělé explozivy si svou neznělost v obecném trendu držely daleko lépe než znělé svou znělost, ani jedna neznělá exploziva nedosáhla tak jednostranného výsledku jako znělá exploziva [b]. Přitom zvýšený počet špatně vnímaných neznělých exploziv často odkryla respondenta, který musel být z nějakého důvodu z výsledků vyloučen. Proto by šlo předpokládat, že takto percepčně silná bude i některá z neznělých exploziv. Tento trend by bylo zajímavé ověřit vyšším počtem respondentů. Buď by došlo k setření tohoto rozdílu, což je pravděpodobnější, nebo naopak k jeho posílení.

Krom návrhu zmíněného v předchozím odstavci lze jako další možné pokračování této práce navrhnout výzkum zaměřený na hlubší prozkoumání stability znělosti při zrychleném tempu artikulace. Vycházet by se přitom mělo z širšího pozorování přirozeného projevu se zvýšeným tempem tak, aby bylo možné lépe stanovit relevantní hodnoty parametrů pro syntézu. Další výzkum by se nabízel také ve srovnání syntézy znělostních charakteristik českých exploziv a frikativ. Zdá se, že byť jsou české frikativy náročnější na fyziologické udržení znělosti než explozivy, tak jako kompenzační mechanismus náročnosti artikulace u nich stačí slabší tónová složka, aby byly ještě vnímané jako znělé. Navrhovaný výzkum by

se mohl zaměřit právě na srovnání tohoto jevu. Jako poslední možný výzkum, který se okrajově odkryl při vyřazování nevhodných respondentů, by mohlo být srovnání vnímané znělosti respondenty v návaznosti na zhoršené kvalitě jejich sluchového vnímání. Všichni tři respondenti, kteří byli z výsledků vyřazeni, byť každý z jiného důvodu, měli převládající tendenci hodnotit explozivy jako znělé, a to i v případech exploziv neznělých. Takový výzkum by mohl odhalit, z jakých důvodů lidé vnímají znělost i v případě, že ve zvukovém signálu znělost přítomna není.

9. Seznam použité literatury

Apple Inc. (2. 6 2013). *Learn more about Siri*. Načteno z Apple: http://www.apple.com/ios/siri/siri-faq/

Beskow, J. (2. 6 2013). *Formant Synthesis Demo.* Načteno z Royal Institute of Technology, Speech, Music and Hearing: http://www.speech.kth.se/wavesurfer/formant/

Biewener, A. A. (2003). *Animal Locomotion.* New York: Oxford University Press.

Boersma, P., & Weenink, D. (nedatováno). *Praat: doing phonetics by computer (v. 5.3.51 x64).* Načteno z Praat: http://www.fon.hum.uva.nl/praat/download_win.html

Dutoit, T. (2001). *An Introduction to Text-to-Speech Synthesis.* Dordrecht: Kluwer Academic Publishers.

Heid, S., & Hawkins, S. (1998). PROCSY: A Hybrid Approach to High-quality Formant Synthesis using HLSyn. *Third ESCA/COCOSDA Workshop on Speech Synthesis*, (stránky 219-224). Blue Mountains, Australia.

Holmes, J., & Holmes, W. (2001). *Speech Synthesis and Recognition* (Sv. 2). London: Taylor & Francis.

Kopeček, I. (2. 6 2013). *Počítačové zpracování řeči, dialogové systémy a asistivní technologie*. Načteno z Fakulta informatiky Masarykovy unierzity: http://www.fi.muni.cz/research/nlp/dialogues.xhtml.cs

Králíček, P. (2004). *Úvod do speciální neurofyziologie.* Praha: Nakladatelství Karolinum.

Linc, R., & Doubková, A. (2003). *Anatomie hybnosti III.* Praha: Nakladatelství Karolinum.

Machač, P. (2006). *Temporální a spektrální struktura českých explozív. Nepublikovaná disertační práce.* Praha: Fonetický ústav FF UK.

Matoušek, J. (2010). Automatic Segmentatio of Parasitic Sounds in Speech Corpora for TTS Synthesis. V S. P., A. Horák, I. Kopeček, & I. Pála, *Text, Speech and Dialogue: 13th International Conference, TSD 2010, Brno, Czech Republic, September 6-10, 2010.Proceedings* (stránky 369-376). Brno: Springer.

Matoušek, J. (2010). Studijní materiál ke kurzu.

O'Regan, G. (2012). *A Brief History of Computing* (2. vyd.). London: Springer.

Palková, Z. (1994). *Fonetika a fonologie češtiny.* Olomouc: Karolinum.

Rojas, R., & Hashagen, U. (2002). *The First Computers.* USA: MIT Press.

Sensimetrics Corporation. (1995). *Excercises from 1995 Summer Course.* Mariefred, Sweden: Sensimetrics Corporation.

Sensimetrics Corporation. (2004). *High-Level Speech Synthesizer User Interface Manual.* Somerville: Sensimetrics Corporation.

Schwarz, D. (2000). A System for Data-Driven Concatenative Sound Synthesis. *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*, DAFX-1 - DAFX-6.

Skarnitzl, R. (2011). *Znělostní kontrast nejen v češtině.* Praha: Nakladatelství Epocha.

Skarnitzl, R., & Volín, J. (Duben 2012). Referenční hodnoty vokalických formantů pro mladé dospělé mluvčí standardní češtiny. *Akustické listy, 18*(1), stránky 7-11.

Sluijter, A., Bosgoed, E., Kerkhoff, J., Meier, E., Rietveld, T., Sanderman, A., . . . Terken, J. (1998). Evaluation of speech synthesis systems for Dutch in telecommunication applications. *Proceedings of the ESCA/COCOSDA International Workshop on Speech Synthesis*, 213-218.

Styger, T., & Keller, E. (1994). Formant Synthesis. V E. Keller, *Fundamentals of Speech Synthesis: Recognition: Basic Concepts, State of the Art, and Future Challenges* (stránky 109-128). Chichester: John Wiley.

Svirsky, M. A., Stevens, K. N., Matthies, M. L., Manzella, J., Perkell, J. S., & Wilhelms-Tricarico, R. (1997). Tongue surface displacement during bilabial stops. *Journal of the Acoustical Society of America, 102*(1), 562-571.

Syntrillium Software Corporation. (2003). Cool Edit Po v 2.1. Phoenix, USA.

Tatham, M., & Morton, K. (2005). *Developments in Speech Synthesis*. Chichester: John Wiley & Sons Ltd.

Troja, S., Druga, R., & Pfeiffer, J. (1990). *Centrální mechanismy řízení motoriky - teorie, poruchy a léčebná rehabilitace.* Praha: Avicenum.

Účinnost (fyzika). (22. 3 2013). Načteno z Wikipedie. Otevřená encyklopedie.: http://cs.wikipedia.org/w/index.php?title=%C3%9A%C4%8Dinnost_(fyzika)&oldid=9984928

Urban, P. (2. 6 2013). *Microsoft chystá zásadní inovaci. Excel porozumí „lidským příkazům“*. Načteno z Cnews: http://extrawindows.cnews.cz/microsoft-chysta-zasadni-inovaci-excel-porozumi-lidskym-prikazum

Václavík, L. (2. 6 2013). *Graph Search: Facebook má vyhledávač, který mu může Google závidět*. Načteno z Cnews: http://www.cnews.cz/clanky/graph-search-facebook-ma-vyhledavac-ktery-mu-muze-google-zavidet

Příloha A

Zácvičné položky:

1. **P** - něco mezi - B
2. P - něco mezi - **B**
3. P - **něco mezi** - B

Iniciály: _____

Pohlaví: _____

Rok narození: _____

Poslechový test:

1. P - něco mezi - B
2. T - něco mezi - D
3. K - něco mezi - G
4. P - něco mezi - B
5. T - něco mezi - D
6. K - něco mezi - G
7. P - něco mezi - B
8. K - něco mezi - G
9. T - něco mezi - D
10. P - něco mezi - B
11. T - něco mezi - D
12. P - něco mezi - B
13. T - něco mezi - D
14. P - něco mezi - B
15. K - něco mezi - G
16. P - něco mezi - B
17. K - něco mezi - G
18. T - něco mezi - D
19. P - něco mezi - B
20. K - něco mezi - G
21. P - něco mezi - B
22. T - něco mezi - D
23. P - něco mezi - B
24. K - něco mezi - G
25. T - něco mezi - D
26. K - něco mezi - G
27. T - něco mezi - D
28. P - něco mezi - B
29. T - něco mezi - D
30. K - něco mezi - G
31. P - něco mezi - B
32. K - něco mezi - G
33. T - něco mezi - D
34. K - něco mezi - G
35. T - něco mezi - D
36. K - něco mezi - G

Příloha B

#	Kvazi-slovo	ag (mm^2)	ue (cm^3)	t (ms)	Popis sledované explozivy
1	apa	30	0	-	exploziva neznělá
2	aba	4	150	-	exploziva plně znělá (nadhrtanové dutiny svůj objem při artikulaci zvětšují)
3	aba (ue)	4	0	-	exploziva znělá, artikulace bez zvětšení nadhrtanových dutin
4	aba (ag)	10	150		exploziva znělá, artikulace s nižším podílem tónové složky
5	aba (t)	4	150	-30	exploziva znělá se zkrácenou tenzní částí
6	aba (t+ue)	4	0	-30	exploziva znělá se zkrácenou tenzní částí, artikulace bez zvětšení objemu nadhrtanových dutin
7	ata	30	0	-	exploziva neznělá
8	ada	4	150	-	exploziva plně znělá (nadhrtanové dutiny svůj objem při artikulaci zvětšují)
9	ada (ue)	4	0	-	exploziva znělá, artikulace bez zvětšení nadhrtanových dutin
10	ada (ag)	10	150		exploziva znělá, artikulace s nižším podílem tónové složky
11	ada (t)	4	150	-30	exploziva znělá se zkrácenou tenzní částí
12	ada (t+ue)	4	0	-30	exploziva znělá se zkrácenou tenzní částí, artikulace bez zvětšení objemu nadhrtanových dutin
13	aka	30	0	-	exploziva neznělá
14	aga	4	150	-	exploziva plně znělá (nadhrtanové dutiny svůj objem při artikulaci zvětšují)
15	aga (ue)	4	0	-	exploziva znělá, artikulace bez zvětšení nadhrtanových dutin
16	aga (ag)	10	150		exploziva znělá, artikulace s nižším podílem tónové složky
17	aga (t)	4	150	-30	exploziva znělá se zkrácenou tenzní částí