



FILOZOFICKÁ FAKULTA  
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Adléta Hanzlová

Realizace a vnímání melodie v šepotu

Fonetický ústav

Vedoucí práce: Ing. Tomáš Bořil, Ph.D.

Studijní program: humanitní vědy

Studijní obor: AAA FON

Praha 2022

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

V Praze dne .....

Podpis

Děkuji především Ing. Tomáši Bořilovi, Ph.D. za veškerou pomoc, ochotu při konzultacích a cenné rady při zpracování této bakalářské práce. Dále bych chtěla poděkovat všem, kteří mi pomohli s dalšími aspekty této práce. Samozřejmě děkuji také mluvčím, které se účastnily nahrávání, a všem účastníkům obou percepčních testů. V neposlední řadě bych ráda poděkovala své rodině za podporu a trpělivost.

**Název práce:** Realizace a vnímání melodie v šepotu

**Autor:** Adléta Hanzřlová

**Ústav:** Fonetický ústav

**Vedoucí bakalářské práce:** Ing. Tomáš Bořil, Ph.D.

**Abstrakt:**

Melodie patří mezi suprasegmentální vlastnosti řeči a její vnímání se odvíjí především od rychlosti kmitání hlasivek, které se projevuje v základní frekvenci ( $f_0$ ). Šepot je však definován absencí fonace a tudíž v něm základní frekvence nutně chybí. Přesto se zdá, že zamýšlená melodie je i v šepotu rozpoznatelná.

V rámci své bakalářské práce se věnuji této problematice z percepčního a akustického hlediska. Představuji percepční test obsahující nahrávky šeptané řeči a zpěvu v šepotu, který prokázal, že rozeznání melodie v šepotu není vždy dílem náhody. Dále pak shrnuji provedenou akustickou analýzu zkoumající vliv zamýšlené melodie v šepotu na pohyb formantů, poměry formantů, spektrální těžiště a spektrální sklon. V této akustické analýze se jako možné koreláty melodie šeptané řeči ukazují F2 a spektrální těžiště signálu bez hlavního formantového pásma. U zpěvu v šepotu potom F2, F3, poměry F2:F1 a F3:F2, spektrální těžiště a spektrální sklon.

**Klíčová slova:** šepot, intonace, absence fonace, základní frekvence, akustické koreláty melodie, percepce

**Title:** Production and perception of melody in whisper

**Author:** Adléta Hanzlová

**Institute:** Institute of Phonetics

**Supervisor:** Ing. Tomáš Bořil, Ph.D.

**Abstract:**

Melody is a suprasegmental feature of speech and its perception depends mainly on the speed of vocal fold oscillation reflected in the fundamental frequency ( $f_0$ ). Whisper is defined by the absence of phonation and therefore the lack of fundamental frequency. Intended melody in whisper, however, seems to be discernable regardless of this lack of  $f_0$ .

In my thesis, I consider the topic of melody in whisper from a perceptual and acoustic point of view. I present a perception experiment assessing the discernability of melody in whispered words as well as words sung in whisper. This experiment proved that melody in whisper in certain cases can in fact be discerned.

I then further assess the effect of intended melody in whisper on formant frequencies, formant to formant ratios, center of gravity and spectral slope. In whispered speech, the acoustical parameters affected by intended melody turned out to be F2 and center of gravity of stop-band filtered signal with frequencies containing main formant bandwidths removed. In words sung in whisper, the affected parameters are F2, F3, F2:F1 and F3:F2 ratios, center of gravity and spectral slope.

**Keywords:** whisper, intonation, absence of phonation, fundamental frequency, acoustic correlates of melody, perception

# Obsah

<b>1</b>	<b>Úvod</b>	<b>8</b>
1.1	Melodie řeči, její vlastnosti a funkce . . . . .	8
1.2	Některé aspekty komunikace v šeptané řeči . . . . .	9
1.3	Výzkum melodie v šepotu . . . . .	10
1.4	Větná intonace v šepotu . . . . .	13
1.5	Pojmosloví . . . . .	15
<b>2</b>	<b>Metoda</b>	<b>16</b>
2.1	Nahrávky . . . . .	16
2.2	Percepční test . . . . .	19
2.3	Akustická analýza . . . . .	22
<b>3</b>	<b>Výsledky</b>	<b>26</b>
3.1	Percepční test . . . . .	26
3.1.1	Úspěšnost v závislosti na zamýšlené melodii . . . . .	26
3.1.2	Úspěšnost v závislosti na zamýšleném intervalu . . . . .	27
3.1.3	Úspěšnost v závislosti na mluvčích . . . . .	30
3.1.4	Zkušenost s hudbou a rozpoznání melodie v šepotu . . . . .	31
3.2	Akustická analýza šepotu . . . . .	31
3.2.1	F1-F3 . . . . .	31
3.2.2	Poměry jednotlivých formantů . . . . .	35
3.2.3	Spektrální těžiště . . . . .	36
3.2.4	Spektrální sklon . . . . .	38
3.3	Akustické koreláty a percepční test . . . . .	39
3.4	Chování parametrů v modální fonaci . . . . .	40
<b>4</b>	<b>Diskuse</b>	<b>41</b>
4.1	Akustické koreláty melodie v šepotu . . . . .	42

4.1.1	Formanty a poměry formantů . . . . .	42
4.1.2	Spektrální těžiště . . . . .	44
4.1.3	Spektrální sklon . . . . .	45
4.2	Zpracování percepčního testu . . . . .	46
4.2.1	Výsledky percepčního testu . . . . .	46
4.2.2	Percepční test v online prostředí . . . . .	47
4.2.3	Přesah možného výzkumu získaných dat . . . . .	48
4.3	Další možnosti směřování výzkumu . . . . .	48
<b>5</b>	<b>Závěr</b>	<b>51</b>
	<b>Seznam použité literatury</b>	<b>53</b>

# 1. Úvod

## 1.1 Melodie řeči, její vlastnosti a funkce

Melodie patří mezi suprasegmentální vlastnosti řeči a odvíjí se především od základní frekvence ( $f_0$ ), která odpovídá rychlosti kmitání hlasivek (Skarnitzl a kol., 2016, str. 125). Čím rychleji hlasivky kmitají, tím je hodnota  $f_0$  vyšší, a tím vyšší tón vnímáme (Ashby a Maidment, 2016, str. 130). Ačkoli je  $f_0$  odpovědná za vnímanou výšku, není přímo úměrná tomu, co považujeme za melodii řeči. Melodie řeči, též intonace, je mentální interpolací průběhu základní frekvence. Drobné odchylky a neznělé úseky řeči (především neznělé konsonanty) nevnímáme jako narušení melodické kontury, ale pomyslně si konturu vyhladíme a vnímáme intonaci jako kontinuální. Zásadní pro rozlišení melodie jsou hodnoty  $f_0$  v jádrech slabik, která bývají nejčastěji tvořena vokály. U vokálů s trváním do 100 ms vnímáme jako jejich frekvenci průměrnou hodnotu  $f_0$  z druhé třetiny průběhu. U delších potom její pohyb (směr) uvnitř prostřední třetiny (Skarnitzl a kol., 2016, str. 127-128). Vnímání intonace je ovšem velmi komplexní a existují i pokročilé modely, které zpracovávají různé parametry a berou v potaz i konkrétní situaci jakou je například jednorázový či opakovaný poslech (Mertens, 2004).

Melodie má v řeči různé funkce, z nichž hlavní jsou funkce lexikální, gramatická, diskurzní, afektivní a indexová. Lexikální funkci má melodie v tónových jazycích, kde slouží k rozlišení významu a existují minimální páry slov lišících se pouze melodickým průběhem. Gramatická funkce se pak projevuje ve všech jazycích, kde melodie řeči (intonace) může odlišovat význam věty (v češtině např. odlišení otázky od věty oznamovací se stejným slovosledem), zároveň také naznačuje syntaktickou a sémantickou strukturu věty, což značně usnadňuje porozumění. Diskurzní funkce je velmi přirozená – často si to ani neuvědomujeme, ale i díky melodii například poznáme, zda mluvčí dokončil promluvu, nebo se chystá dále pokračovat, což zajišťuje plynulost konverzace a předávání slova. Jelikož ob-



vykle neprodukuje plochou, robotickou řeč, či to ani není v našich silách, naše projevy nejsou nikdy zcela neutrální, a různé dlouhodobé i krátkodobé stavy se promítají v naší intonaci. Tato funkce melodie řeči je označována jako funkce afektivní. Indexová funkce pak značí vztah mezi melodií řeči a identitou mluvčího, jeho zařazení do sociální skupiny (Skarnitzl a kol., 2016, str. 128-133).

Stanovili jsme tedy, že intonace závisí na základní frekvenci hlasivek a je pro řečovou komunikaci velmi zásadní, či dokonce nezbytná. Zajímavá situace však nastává, když se zamyslíme nad komunikací v šepotu. Šepot je definován absencí fonace – hlasivky nekmitají, a tak neprodukují zvuk, který by bylo možné označit za hudební tón (Ashby a Maidment, 2016, str. 31). Zároveň ale hlasivky nejsou zcela otevřené jako při běžném dýchání. Hlasivkové vazy jsou přiblížené (až sevřené, viz Pompino-Marschall, 2003, str. 35) a mezi hlasivkovými chrupavkami je trojúhelníkovitá štěrbina, kterou prochází vzduch, čímž vzniká šum. Na rozdíl od zvuku produkovaného (kvaziperiodickým) kmitáním hlasivek je jednou z vlastností šumu nepravidelný průběh (Bořil, 2017), a tedy nelze určit jeho základní frekvenci. Proto například počítačový program Praat (Boersma a Weenink, 1992–2021) nenalezne v šepotu základní frekvenci pomocí postupů, které využívá k určování základní frekvence běžné řeči. Z těchto vlastností šepotu vyvstávají dvě zásadní otázky. Je-li melodie řeči zásadním prvkem komunikace, rozeznáme zamýšlenou intonaci i v šepotu? Budeme-li považovat šepot za formu plnohodnotné komunikace, logickým závěrem by bylo, že melodie rozeznatelná bude. A konečně, je-li melodie v šepotu identifikovatelná, jaká vlastnost nám umožňuje ji rozeznat, když základní frekvence kmitání hlasivek chybí a hlasivkové buzení nahrazuje šum?

## 1.2 Některé aspekty komunikace v šeptané řeči

Šepot jako forma komunikace bývá využíván jak na veřejnosti, tak i v domácím prostředí. Častými důvody k šepotu ve veřejném prostoru jsou například snaha nevyrušovat okolí, předat důvěrnou informaci konkrétnímu člověku nebo potvrdit

blízký vztah k druhé osobě. V soukromí pak může být šepot využíván k vyjádření náklonnosti nebo jako výzva k vzájemné interakci (Cirillo, 2004). Další využití má šepot například v hlasových a hereckých cvičeních jako prostředek umožňující soustředit se plně na dech a pohyb svalů v těle, aniž by činnost hlasivek odváděla pozornost (viz Puffer, 2018, str. 43, 50).

Šepot má ve srovnání s modální fonací ve spektru nižší amplitudy v nižších frekvencích (Wennndt a kol., 2002). Zároveň má nižší intenzitu zvuku a je znatelně tišší (Zhang a Hansen, 2007). V šepotu také obecně mluvíme pomaleji a trvání všech hlásek je delší než u modální fonace (Bonnot a Chevrie-Muller, 1991). I přes nepřítomnost hlasivkového tónu v šeptaných vokálech stále nalézáme formanty, které jsou zásadní pro rozlišení kvality vokálu (Romportl, 1973, str. 42). Romportl ve svých *Základech fonetiky* (1973, str. 144) dokonce zmiňuje využití šeptaných vokálů k určování formantů jako vhodný postup při poslechové analýze. Určování formantů poslechem jako takové však v dnešní době spíše záhodno není (navíc se později ukázalo, že formanty v šepotu jsou vůči modální fonaci obecně vyšší, viz Ito a kol., 2005).

Ač je nám šepot v běžné komunikaci srozumitelný, k rozpoznání šeptané řeči v hláskové i intonační rovině nemusí nutně sloužit vlastnosti, které jsou prominentní v běžné fonaci. Příkladem tohoto jevu je i značně snížená funkčnost automatického rozpoznávače řeči, natrénovaného na běžné fonaci, při pokusu rozpoznat šeptanou řeč (Ito a kol., 2005). Rozeznání melodie v šepotu můžou umožňovat (sekundární) koreláty v obdobné míře přítomné i v běžné řeči, také se však v šepotu mohou silněji uplatňovat jiné prvky (kompenzační), které v běžné řeči vnímání melodie tolik neovlivňují (Heeren a van Heuven, 2014).

### 1.3 Výzkum melodie v šepotu

Jeden z prvních zaznamenaných experimentů v souvislosti s výzkumem šepotu provedl v roce 1830 Déleau, který ověřil, že řeč může vznikat i bez činnosti hlasivek, když v rámci pokusu zavedl elastickou trubici skrz nos do hltanu, načech

uzavřel glotis, do trubice pustil vzduch a zkusil artikulovat. Takto vzniklá sekvence segmentů byla dostatečně srozumitelná, aby bylo možné ji považovat za řeč – vytvořil tak jakýsi umělý šepot. (Panconcelli-Calzia, 1994, str. 35-36)

Donders (1858, str. 158) později využil šepot k určení rezonance vokálů (formantů) podle nejsilnější složky identifikovatelné v šepotu. Vycházel z předpokladu, že ústní dutina umocní rezonance proudu vzduchu procházející přes hlasivky nastavené na šepot. Vokál vždy v hlasitějším šepotu dlouze držel a následně určil hudební tón (frekvenci) vyprodukovaného šumu. U vysokých a středových zaokrouhlených vokálů navíc při dostatečně silném proudu vzduchu přešel až do pískání, podle kterého rezonanci určil. Srovnání vnímané frekvence šepotu s čistým tónem a přechod z šepotu do pískání potom jako jednu z metod v rámci rozsáhlé studie rezonance vokálů v závislosti na kvalitě využil i Helmholtz (1863, str. 171-172), který se neřídil pouze sluchem, ale rezonance určoval pomocí rezonátorů a ladiček, čímž mohl dojít k přesnějším výsledkům než Donders. Helmholtzův výzkum je známý a položil základy pozdějšímu zkoumání formantů. Užití šepotu k těmto účelům je však méně známá metoda, která zvláště s technologickým pokrokem příliš nezískala na popularitě – opírá se totiž zčásti o poslechovou analýzu, která nemusí být jednoznačně objektivní, a tudíž není vždy preferována. Jak ale poznamenává I. B. Thomas (Thomas, 1969), Helmholtzovy experimenty s šeptanými vokály poukazují na velmi jasnou spojitost mezi vnímanou výškou (šeptaných) vokálů a frekvencí jejich formantů, a tedy představují první náznaky výzkumu projevů melodie v šepotu. Toto téma však nebylo v době počátků fonetiky blíže příliš zkoumáno.

Větší debata napříč akademickým prostředím ohledně melodie v šepotu se odehrála až v návaznosti na studii, ve které Panconcelli-Calzia (1954) důrazně popřel možnost, že by melodie v šepotu mohla být rozeznatelná, a zpochybnil Gietovy (1950, str. 95-97) výroky ohledně srozumitelnosti šeptaných tónových jazyků. Giet (1955) na tuto studii reagoval a argumentoval tím, že nejen v tónových jazycích, kde je melodie zásadní k rozlišení významu, je prozodie neod-

dělitelnou součástí řeči, která napomáhá srozumitelnosti a přirozenosti sdělení. U šepotu jako plnohodnotného způsobu komunikace musí tedy existovat mechanismus, který kompenzuje absenci fonace, aby byl zachován komunikační záměr.

Meyer-Eppler (1957) navázal na Gieta experimentem, kde nahrával německá slova se dvěma melodickými průběhy a vokály zpívané v šepotu. Spektrografickou analýzou došel k závěru, že u hlásek [i, e, a] pohyby základní frekvence nahrazuje spektrální šum (u vyšších tónů je ve vyšších frekvencích více šumu a dochází k rozostření vyšších formantů), zatímco u hlásek [a, u] navíc dochází i k posunu formantů. Výsledky byly stejné u slov i u zpívaných vokálů. Pozorováním navíc zjistil, že při šeptaném zpěvu vyšších tónů se zvedá larynx a tedy nejspíše dochází ke zúžení hlasivkové štěrbiny. (Pohyby hrtanu byly později přesněji změřeny. Přesto, že není pohyb tak výrazný jako při modální fonaci, je stále nezanedbatelný, viz Coleman a kol., 2002).

Kloster Jensen (1958) pak ověřoval srozumitelnost šeptaných minimálních párů v tónových jazycích pomocí několika percepčních testů. U různých jazyků se výsledky lišily, shledal však, že míra porozumění převyšuje pouhé hádání. Jeho závěry však (nejen) v dnešním světle nelze považovat za zcela spolehlivé, protože percepční testy zahrnovaly někdy jen jednoho respondenta, který v případě švédštiny byl zároveň i nahraným mluvčím (nahrávky vlastní řeči identifikoval se 100% úspěšností). Výzkum šeptaných tónových jazyků a debata s ním související dále pokračovaly (navazuje např. Segerbäck, 1965), postupně se však v návaznosti na Meyer-Epplerovu spektrografickou analýzu začaly objevovat také studie zabývající se melodií v šepotu nejen jako lexikální jednotkou, ale i intonací a projevy melodie v jazycích, které nejsou považovány za tónové.

S odkazem na Helmholtzův výzkum formantů s využitím šepotu ověřoval I. B. Thomas (1969) vztah mezi formanty a vnímanou výškou vokálu. Z nahrávek anglických vokálů vytvořil vždy smyčku, kde se jeden vokál nepřetržitě opakoval. Posluchači (s hudebním vzděláním) poté ladili zdroj čistého tónu na stejný tón, jaký slyšeli v nahrávce, která jim byla prezentována pouze jako šum, nikoli šepot.

Tón, který posluchači naladili, odpovídal hodnotě F2 získané ze spektrogramu. Jelikož se posluchači ve svých odhadech shodovali, zopakoval Thomas stejný experiment ještě jednou, tentokrát však upozornil posluchače, že v šumu mohou slyšet i více tónů. U zadních nízkých vokálů tak navíc získal tón, který odpovídal hodnotě F1. Svým výzkumem tedy nevyvrátil účinnost Helmholtzových postupů při určování frekvencí formantů. Navíc také podotýká, že posluchači často tón ladili podle vlastního pískání, aniž by k tomu byli jakkoli vedeni.

Jak Thomas uvádí, jeho výsledky se neshodují se závěry, ke kterým předtím došel Meyer-Eppler (1957), domnívám se však, že charakter každého z těchto experimentů je mírně odlišný – zatímco Thomas zkoumal, jakou frekvenci posluchač v šeptaném vokálu vnímá jako dominantní, Meyer-Eppler se snažil nalézt korelát, který by byl odpovědný nikoli za absolutní vnímanou výšku, ale za melodický pohyb uvnitř slova, tedy výšku relativní, ve vztahu k okolí. Proto také srovnával, jak se v závislosti na změnách zamýšlené výšky vokálu mění různé parametry.

## 1.4 Větná intonace v šepotu

Technologický pokrok v posledních desetiletích značně usnadnil zpracování řeči pro účely fonetického výzkumu. Vývoj počítačů a s ním spojený vznik fonetických počítačových softwarů v mnohém zjednodušil například spektrální analýzu, která byla dříve výrazně náročnější (Boersma a Weenink, 2001). Velmi rozšířeným fonetickým počítačovým programem je Praat (Boersma a Weenink, 1992–2021), který umožňuje nesčetné množství analýz a manipulací. Díky těmto technologiím se možnosti experimentální fonetiky rozšířily (Boersma a Weenink, 2001). S pokrokem souvisí i postupné zaměření pozdějších výzkumů melodie v šepotu na různé spektrální vlastnosti v návaznosti na Meyer-Epplerovy (1957) postupy.

Detailnější studie zabývající se percepcí a akustickými korelátly melodie v šepotu ukázaly, že i když jsou některé vlastnosti modální fonace a šepotu obdobné, posluchač nejspíše rozpoznává melodii v šepotu na základě jiných parametrů než u běžné fonace. Zatímco u modální fonace hraje zásadní roli  $f_0$  a nižší frekvenční

pásma, u šepotu se zásadní informace pro odlišení melodie nachází ve frekvencích nad 1,5–2 kHz (Heeren a Lorenzi, 2014).

Dřívější analýzy potvrdily, že pohyb formantů odpovídá zamýšlené intonaci (Fónagy, 1969; Higashikawa a Minifie, 1999). Percepční testy s různě filtrovanými šeptanými promluvy (Heeren a Lorenzi, 2014) a analýza syntézou (Higashikawa a Minifie, 1999) však ukázaly, že formanty samotné nemusí být nutně hlavním vodítkem pro rozeznání melodie v šepotu. Odstranění pásma nižších frekvencí obsahujícího F1 nemělo vliv na rozpoznatelnost melodie, odfiltrování pouze F2 nebo F3 také nezhoršilo výsledky – pokud tedy pohyb formantů vnímáme jako pohyb melodie, musí jít o pohyb několika formantů zároveň a jejich vzájemnou interakci (Heeren a Lorenzi, 2014). Toto zjištění nevyvrací dřívější závěry studie využívající syntézu šeptaných vokálů, ve které posun F1 a F2 zároveň dosahoval v percepčním testu lepší rozpoznatelnosti než posun pouze jednoho z těchto formantů (Higashikawa a Minifie, 1999). Kromě formantů nejspíše důležitou roli v rozpoznání melodie v šepotu hrají také spektrální sklon (Heeren a Lorenzi, 2014) a spektrální těžiště, které se v šepotu v závislosti na zamýšlené výšce mění znatelněji než v modální fonaci (Heeren, 2015).

Studie z posledních let zabývající se melodií v šepotu se z evropských jazyků zaměřily na produkci a percepci melodie ve francouzských (Heeren, 2015) a holandských (Heeren a van Heuven, 2014) slovech a dále v samostatných slabikách (pseudoslovech) čtených rodilými mluvčími holandštiny (Heeren, 2015). Nejnovější výzkum pak směřuje zejména k syntéze a automatickému rozpoznání šepotu a jejich užití například v zařízení chytré domácnosti (Raeesy a kol., 2018; Phapatanaburi a kol., 2022), v této oblasti je dominantním jazykem především angličtina. Šepot v češtině byl zatím zkoumán v kontextu realizace znělosti (Machač a Šturm, 2010; Skarnitzl a kol., 2013; Svatošová, 2021). V této bakalářské práci se proto zaměřuji na realizaci a vnímání melodie v šepotu v českém prostředí, s užitím českých dvouslabičných slov. První část tvoří percepční test, který je zaměřen na rozeznatelnost melodie v šepotu v různě vzdálených (hudebních)

intervalech. Ve druhé části se zabývám akustickou analýzou různých korelátů, které se z dosavadních výzkumů jeví být nosnými prvky intonace, jako je spektrální těžiště, spektrální sklon a také formanty, které jsou s melodií v šepotu spojovány nejdéle.

## 1.5 Pojmosloví

Pojem *melodie* v této práci užívám ve dvou významech – *melodie řeči* odpovídá intonaci (ve smyslu změny melodického průběhu), zatímco *melodie* v souvislosti se zpěvem a zpěvem v šepotu označuje melodii hudební, tedy sled (hudebních) tónů. Vzhledem k charakteru experimentu popsaného v následujících kapitolách však ne vždy tyto dvě interpretace pojmu striktně odlišuji. Pro užití pojmu *melodie* pro oba významy, raději než možná očividnějšího a snáze odlišitelného pojmu *intonace* pro melodii řeči, jsem se rozhodla právě proto, že se obě *melodie* vzájemně zčásti překrývají, jelikož je materiál, který používám, založen na hudební stupnici a frekvencích hudebních tónů i v případě řeči. Pojem *tón* používám v následujících kapitolách čistě ve významu hudebního tónu (tedy nikoli např. tónu v tónových jazycích), *interval* pak značí sled dvou tónů či jejich vzdálenost. Pro sjednocení popisu používám pojem *stoupavý interval* a *klesavý interval* a názvy hudebních intervalů u obou typů zkoumaného projevu, tedy řeči i zpěvu. Vzdálenost mezi dvěma tóny v rámci jednoho intervalu udávám v půltónech (*st*), které jsou při výzkumu intonace vhodnou jednotkou (Volín, 2010, str. 41). Vzhledem k tomu, že použité nahrávky vznikly za kontrolovaných podmínek a s cílem realizovat přesný hudební tón, a jsou tedy do jisté míry umělé, nenapodobující spontánní řeč, považuji užití spíše hudebního názvosloví za postup, který usnadní porozumění. Jsem si však vědoma, že pro suprasegmentální řečové jevy jsou zavedeny pojmy zaměřené na jednoznačnější prozodický popis (viz Palková a kol., 2004).

## 2. Metoda

### 2.1 Nahrávky

Pro zkoumání melodie v šepotu jsem nahrávala dva typy dvouslabičných slov, realizovaných na různě velkých intervalech. Prvním typem byly šeptané věty ve formě „Řekl [cílové slovo realizované na požadovaném intervalu] anebo [obdobné slovo]“. Celkem čtyři cílová slova (jéjé, lálá, jójó, bába) byla realizována ve čtyřech různých intervalech (sekunda dolů, tercie nahoru, kvinta dolů, kvinta nahoru). Slovo na konci věty vždy odpovídalo variantě zkoumaného slova s krátkými vokály (jeje, lala, jojo, baba).

Výběr slov k nahrávání se odvíjel od několika parametrů: aby byly vlastnosti jednotlivých slabik co nejlépe vzájemně porovnatelné, jednalo se vždy o dvouslabičné slovo se strukturou CVCV, kdy oba konsonanty a vokály měly shodnou kvalitu. Všechna slova s výjimkou jednoho obsahovala dvě stejné slabiky s dlouhými vokály. Pro zachování (v případě šepotu pomyslného) průběhu  $f_0$  napříč celým slovem jsem vybrala především slova obsahující aproximanty, které v češtině nemají neznělý protějšek. Slovo „bába“ se zmíněným parametrům mírně vymyká, splňuje však velmi dobře poslední podmínku, kterou byl rozeznatelný význam slova (existence daného slova v českém jazyce). Slovo bylo vždy začleněno do věty tak, aby na něm mohl být realizován požadovaný interval, ale zároveň se vyskytlo uprostřed souvislé řeči, čímž se zmenšila šance, že druhá slabika bude ovlivněna některým z jevů, které mohou nastat u finálních hlásek před pauzou, jako je například dyšná či třepená fonace a další vlivy (Machač a Skarnitzl, 2010, str. 137). Za slovem následovala spojka „anebo,“ která se vyslovuje lépe než jiná synonyma (např. „či“) a nezačíná nazálou (jako např. „nebo“), čímž se snižuje pravděpodobnost, že by druhý vokál v cílovém slově byl vlivem koartikulace nazalizovaný.

Druhým typem nahrávek byly čtyři různě velké intervaly (1 st, 2 st, 4 st,



7 st) zpívané v šepotu, každý relizovaný ve stoupavé i klesavé verzi ve dvou variantách – jednou na slabice [la:] a poté na slabice [jo:]. Každý ze dvou tónů byl realizován na samostatné slabice. V tomto případě bylo trvání slabik výrazně delší a navíc regulováno způsobem nahrávání, cílový interval byl proto vždy nahrán samostatně, bez přidaného hláskového okolí.

Sada nahrávek od jedné mluvčí celkem obsahovala nahrávky šestnácti mluvených a šestnácti zpívaných intervalů. Všechny typy promluv jsem nahrála jak v šepotu, tak pro srovnávací účely i v modální fonaci (celkem tedy soubor nahrávek od jedné mluvčí obsahoval 32 položek).

Pro nahrávání tohoto materiálu ke zkoumání melodie v šepotu se jevílo nejspolehlivějším zvolit postup, kdy mluvčí opakuje položku, kterou předtím slyší, což umožnilo do jisté míry kontrolovat produkované melodické kontury. Bylo proto nutné nejdříve vytvořit vzorové nahrávky.

Vzorové věty jsem nahrála v domácím prostředí v tichém pokoji s krátkým dozvukem s použitím kardioidního kondenzátorového mikrofonu a programu Audacity (Audacity Team, 1999–2021). Cílový interval (druhé slovo v každé větě) jsem pak ručně doladila pomocí manipulace v programu Praat (Boersma a Weenink, 1992–2021) se snahou zachovat co nejpřirozenější melodickou konturu (postupem Sound > To manipulation > ruční manipulace cílové kontury > Publish resynthesis). Frekvence v oblasti kolem středu vokálů u zkoumaných slov jsem posunula tak, aby odpovídaly požadovanému tónu. Přesnou frekvenci jsem určila podle referenční tabulky hudebních tónů (Suits, 1998–2022). Přehled intervalů, na kterých byla realizována cílová slova, znázorňuje obrázek 2.1 Jelikož se jedná o absolutní (nikoli relativní) výšku, odvíjející se od hudebních tónů, je pro záznam použit notový zápis.

Pro nahrávání zpívaných intervalů jsem připravila sled tónů se zvukem klavíru v programu GarageBand (Apple Inc, 2004–2022). Každý interval zazněl dvakrát za sebou, jednou jako předloha a jednou společně se zpěvem nahrávané osoby. Všechny intervaly byly pravidelně rozmístěny a doprovázeny zvukem metronomu



Obrázek 2.1: Notový zápis intervalů použitých v rámci předlohy k nahrávání mluvených vět. Záznam je pro snazší orientaci uveden o oktávu výše než reálné předlohy. Pořadí kombinací jednotlivých intervalů a slov bylo pro každou nahrávanou mluvčí náhodné.



Obrázek 2.2: Notový zápis intervalů sloužících jako předloha k nahrávání zpěvu a zpěvu v šepotu. Pořadí jednotlivých intervalů bylo pro každou nahrávanou mluvčí náhodné.

pro snadnou orientaci při zpěvu, což zároveň zajistilo srovnatelné trvání všech nahrávaných slabik. Každý přehrávaný tón trval vždy 1 sekundu. Přehled intervalů pro nahrávání zpěvu ukazuje obrázek 2.2.

Nahrávání se zúčastnily čtyři ženy ve věkovém rozmezí od 20 do 24 let. Všechny mají základní hudební vzdělání a zkušenosti se sborovým či sólovým zpěvem.

Využití připravené předlohy bylo mírně odlišné pro nahrávání vět a zpívaných intervalů. V prvním případě vzorová věta zazněla z reproduktoru ve studiu a mluvčí ji sama zopakovala. Předlohu pro zpívané intervaly mluvčí poslouchaly do sluchátek během nahrávání, vždy nejdříve vyslechly interval, a když zazněl podruhé, zazpívaly ho společně s nahrávkou.

Všechny položky byly pro srovnávací účely nahrány v běžně fonované a šepotané verzi. Běžná fonace byla vždy zařazena jako první, aby se mluvčí seznámily s charakterem daného úkolu nejdříve v tradičnější podobě zahrnující rozeznatelnou melodii. Poté se stejná úloha opakovala, ale mluvčí opakovaly věty a zpívané intervaly v šepotu. Pro nahrávání běžné fonace i šepotu byly použity stejné vzorové položky. Během nahrávání jsem kontrolovala, že mluvčí chápou princip zadaného úkolu, případně jsem je přerušila, upozornila, jak projev přizpůsobit

cílovým požadavkům, a nahrávání daných položek opakovala.

Nahrávání proběhlo v nahrávacím studiu fonetického ústavu FF UK. Účastnice dostaly instrukce k danému úkolu vždy před nahráváním dané části, tj. nejdříve si vyslechly pokyny k nahrávání vět, které následně nahrály, poté byla zařazena krátká přestávka, během které dostaly pokyny k nahrávání zpívaných intervalů, a následovala druhá část nahrávání. Každá účastnice dostala jednotlivé položky v rámci každého úkolu v jiném pořadí vygenerovaném pomocí stránky pro náhodná pořadí Random.org (Haahr, 1998–2018). Pořadí položek zůstalo pro jednu mluvčí stejné u obou typů realizace (modální fonace, šepot), v případě zpívaných intervalů i u obou slabik. Celkem tedy nahrály vždy dvakrát stejnou sekvenci vět a čtyřikrát stejnou sekvenci zpívaných intervalů.

Pro další zpracování jsem z každé nahrávky vždy vystříhla pouze slovo, na kterém byl realizován požadovaný interval. Získala jsem tak od každé mluvčí po 32 nahraných dvouslabičných slovech v modální fonaci, šepotu, zpěvu a zpěvu v šepotu.

## 2.2 Percepční test

Abych určila, zda je melodie v šepotu skutečně rozeznatelná, provedla jsem dva percepční testy na stejném principu – první obsahoval nahrávky šeptaného zpěvu, druhý nahrávky slov vystřižených z šeptaných vět. Oba percepční testy jsem vytvořila pomocí online nástroje PsyToolkit (Stoet, 2017, 2010).

Ze sestříhaných nahrávek šeptaných slov jsem pro použití v percepčním testu vyřadila ty, ve kterých byla slyšet činnost hlasivek, a tedy celé slovo nebo jeho část bylo realizováno spíše tichou dyšnou fonací než šepotem. Tím jsem počet položek omezila na 54 u slov nahraných v rámci vět a 53 u intervalů zpívaných v šepotu. Ke zkoumaným nahrávkám jsem vždy ještě přidala další položky, u kterých jsem se za účelem motivace posluchače snažila zajistit snazší rozpoznatelnost melodie. U zpívaných intervalů jsem jako distrakční položky použila devět jiných nahrávek zpívaných intervalů, kde byl rozdíl mezi tóny výrazný. U slov vystřiže-

ných z vět jsem pak vybrala dvanáct ze zkoumaných nahrávek, kde se mi melodie zdála znatelnější, u kterých jsem následně u nižšího ze dvou tónů vždy odfiltrovala frekvence nad 6300 Hz (přibližná hranice, kde končila formantová struktura pozorovatelná ve spektrogramu). Percepční testy tak měly celkem 62 a 66 položek.

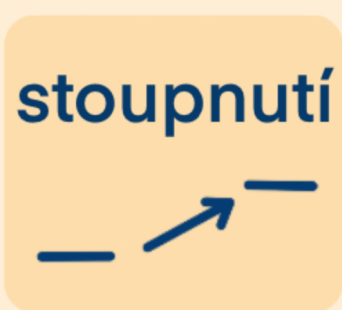
Percepční testy byly provedeny metodou 2AFC (two-alternative forced choice) a jejich zadání bylo u obou formulováno stejně. Respondenti si vždy poslechli nahrávku a měli za úkol určit, zda v melodii slyší stoupnutí, či klesnutí. U každé položky měli možnost si ji jednou poslechnout znovu. Tlačítka „stoupnutí“ a „klesnutí“ byla doplněna názornými obrázky. U intervalů zpívaných v šepotu byla tlačítka umístěna vedle sebe, u slov z vět pod sebou. Umístění označení dvou tlačítek bylo vždy v rámci každého percepčního testu pro každého respondenta programem náhodně vygenerováno. Příklad prostředí percepčního testu ukazuje obrázek 2.3. Abych minimalizovala vliv vizuálních vjemů na vnímání jednotlivých položek percepčního testu, tlačítka se vždy objevila až poté, co položka poprvé zazněla. Během opakovaného poslechu jedné položky už tlačítka zůstala viditelná.

Vzhledem k absenci základní frekvence v šepotu jsem za správnou odpověď v rámci obou percepčních testů považovala shodu mezi odpovědí respondenta a zamýšlenému pohybu  $f_0$ , přičemž zamýšleným pohybem míním reálný pohyb melodické kontury v předloze k nahrávání dané položky.

U obou percepčních testů předcházela samotnému experimentu vždy nejdříve krátká instruktáž, kde se respondenti seznámili s ovládním prostředí, a následně šest zácvičných položek, na kterých si vyzkoušeli poslechový úkol. Před začátkem byli všichni vyzváni, aby se experimentu věnovali v klidné místnosti a se sluchátky. Před každým přehráním položky zaznělo krátké desenzitační pípnutí. Pořadí jednotlivých položek v každém percepčním testu bylo pro každého respondenta programem náhodně vygenerováno. Po přehrání poloviny položek u obou percepčních testů a po dokončení prvního byla zařazena krátká pauza s hudebním doprovodem.

Online percepční test vyplnilo 33 respondentů (32 rodilých mluvčích češtiny

1 / 62



**přehrát  
znovu**

**potvrdit  
volbu**

Obrázek 2.3: Ukázka prostředí percepčního testu obsahujícího nahrávky zpěvu v šepotu se zobrazenou možností „klesnutí“. Pořadí dvou horních tlačítek bylo pro každého respondenta náhodné

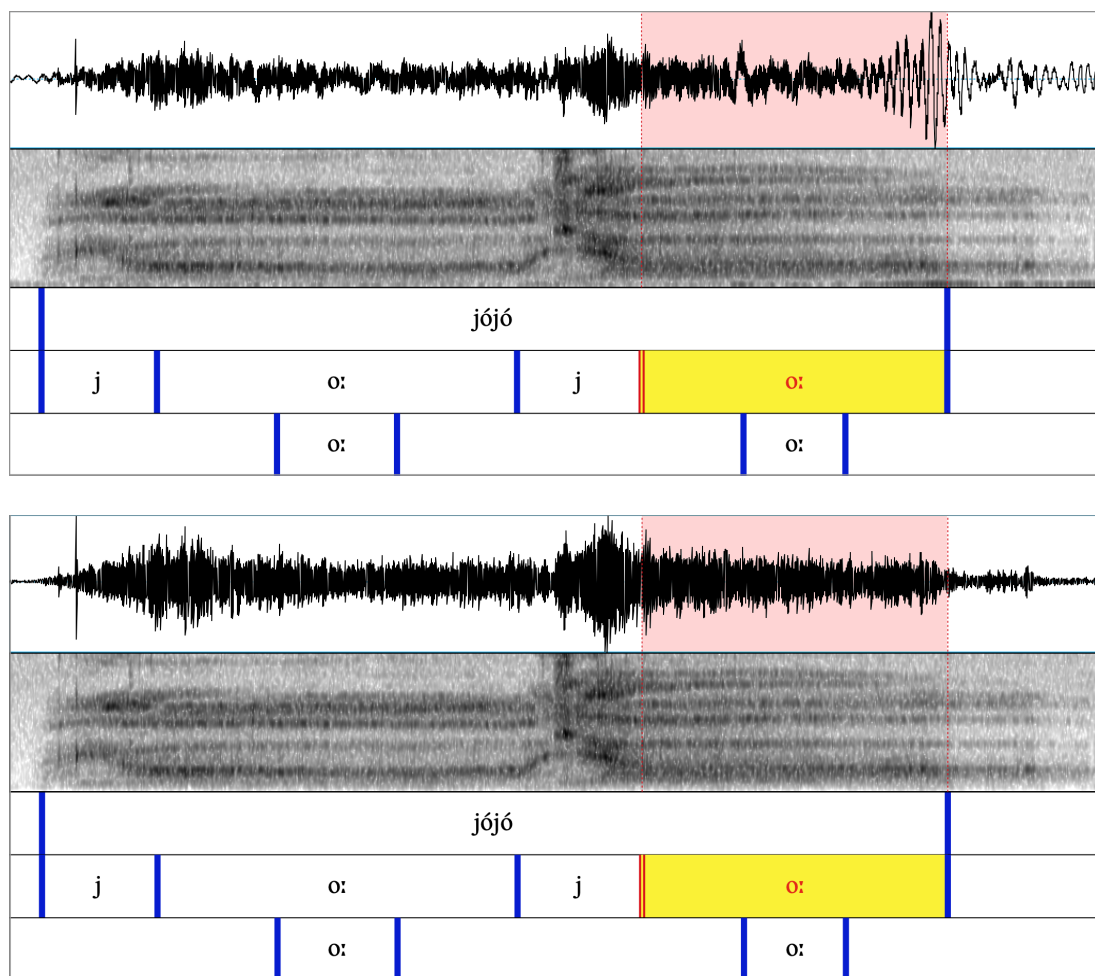
a 1 slovenštiny) ve věku od 17 do 63 let (medián 24), z toho 20 žen a 13 mužů. 23 respondentů uvedlo, že mají zkušenost se zpěvem či hrou na hudební nástroje. Oba percepční testy a krátký dotazník na začátku respondenti dokončili za 13-29 minut (medián 19).

## 2.3 Akustická analýza

Pro účely akustické analýzy jsem ke všem jednotlivým vystřiženým slovům (v šepotu i v modální fonaci) vytvořila textgridy v Praatu (Boersma a Weenink, 1992–2021). Textgrid obsahoval vždy vrstvu s celým slovem a vrstvu hláskové segmentace, kterou jsem nejdříve automaticky nastřelila funkcí „align interval“ a hranice hlásek následně ručně zarovнала podle pravidel popsanych v knize *Fonetická segmentace hlásek* (Machač a Skarnitzl, 2010). Při segmentaci jsem se řídila především spektrogramem a umístění hranic vždy kontrolovala poslechem. Pomocí skriptu v Praatu jsem pak přidala třetí vrstvu, která označila prostřední třetinu vokálů.

V některých nahrávkách šepotu (z tohoto důvodu nevyužitých v percepčním testu) se místy vyskytl náznak činnosti hlasivek (krátký okamžik spíše dyšné fonace než šepotu). Pokud znatelné periody v nižších frekvencích přesahovaly i do prostřední třetiny vokálu, na základě viditelného vrcholu ve spektru daného vokálu jsem filtrem pásmová zádrž odfiltrovala příslušné frekvenční pásmo. Stejně jsem postupovala i pokud v nahrávce byl znatelný ruch v nízkých frekvencích (do 100 Hz) způsobený například průjezdem většího vozidla kolem budovy během nahrávání. Ukázka tohoto postupu je znázorněna na obrázku 2.4.

V prostřední třetině každého vokálu jsem pomocí skriptu v Praatu změřila hodnoty formantů (v Hz), spektrální těžiště (v Hz) a spektrální sklon (v dB). Formanty jsem analyzovala Burgovou metodou mezi 0 a 5500 Hz (výchozí nastavení Praatu pro nahrávku dospělé ženy) a maximálním počtem formantů nastaveným na 5 a následně extrahovala průměrné hodnoty F1-F3. Kromě samotných formantů jsem následně ke každé položce vypočetla poměr všech kombinací těchto



Obrázek 2.4: Slovo „jójó“ zpívané v šepotu s ruchem v nižších frekvencích před filtrací a po filtraci filtrem pásmová zadrž od 0 do 100 Hz. Spektrogram zobrazuje frekvence od 0 do 8 kHz, časové rozlišení je 5 ms.

formantů (tedy F2:F1, F3:F2 a F3:F1). Spektrální těžiště jsem určila ze spektra s použitím výchozího nastavení Praatu. U vokálů ze slov z šeptaných vět jsem navíc změřila spektrální těžiště po filtraci signálu filtrem pásmová zádrž mezi 1000 a 6300 Hz, kde se nacházela hlavní formantová pásma viditelná ve spektrogramu. Spektrální sklon jsem pak určila funkcí „get slope“ z dlouhodobého spektra (LTAS) metodou „energy“ mezi pásmy od 50 Hz do 2 kHz a od 2 do 8 kHz.

Vzhledem k charakteru nahrávaných položek (a jejich regulace formou opakování) jsem se rozhodla do analýzy nezahrnovat parametr trvání vokálů, který je s rozlišením melodie spojován v souvislosti s tónovými jazyky (Liu a Samuel, 2004). Pro rozeznání intonace v šepotu v případě nahrávek, jejichž rytmus může být ovlivněn zadáním k nahrávání, se nezdá být zásadním (Heeren, 2015).

U nahrávek modální fonace jsem za účelem srovnání realizací mluvčích na základě poslechu a předloh použitých k nahrávání změřila i hodnoty  $f_0$  z každého vokálu. Vokály v nahrávkách mluvených slov měly trvání kolem 200-350 ms. U zpívaných slabik bylo trvání výrazně delší, kolem 700-900 ms. V rámci středu těchto vokálů nedocházelo k výraznějším pohybům základní frekvence. Z obou typů nahrávek jsem proto v Praatu určila průměrnou hodnotu  $f_0$  v prostřední třetině vokálu.

Vliv zamýšlené melodie na měřené parametry jsem analyzovala lineárními smíšenými modely v programu R (R Core Team, 2022) s použitím knihoven lme4 (Bates a kol., 2015) a emmeans (Lenth, 2022). Závislou proměnnou byl vždy zkoumaný parametr změřený v prostřední třetině jednotlivých vokálů (1. až 3. formant, podíl formantů, spektrální těžiště, spektrální sklon). Jako fixní faktory jsem použila zamýšlený směr pohybu pomyslné  $f_0$  v rámci daného intervalu (stoupavý/klesavý) v interakci s umístěním konkrétního vokálu v rámci intervalu (v první/druhé slabice cílového slova). Náhodnými faktory pak byly mluvčí a nahrávaná slova. Abych zabránila heteroskedasticitě, hodnoty formantů a spektrálního těžiště jsem pro použití v modelu zlogaritmovala. Statistickou významnost



vlivu fixních faktorů na zkoumané parametry jsem vyhodnotila pomocí testů poměru věrohodností (likelihood ratio tests).

U parametrů, které se ukázaly jako statisticky významné, jsem potom dále zkoumala přímou souvislost podílu těchto parametrů v první a druhé slabice každého slova s úspěšností rozpoznání dané položky v percepčním testu. Tuto souvislost jsem orientačně analyzovala v R pomocí korelačních grafů zobrazených s použitím knihovny ggplot2 (Wickham, 2016).

## 3. Výsledky

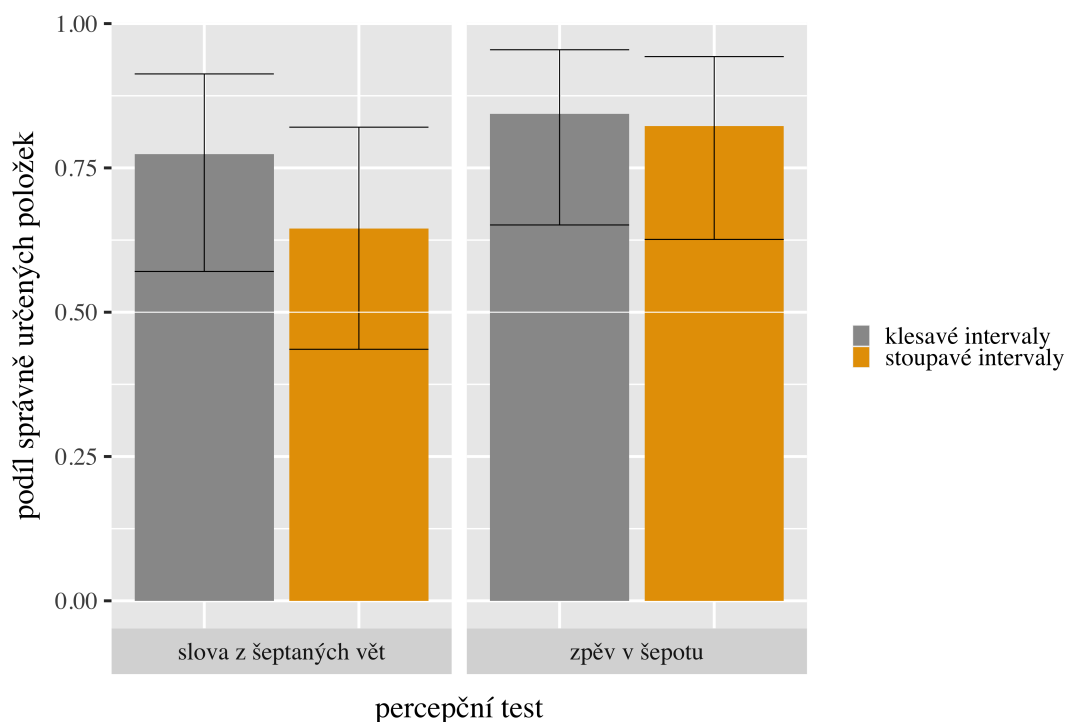
### 3.1 Percepční test

Celková úspěšnost v percepčním testu obsahujícím nahrávky slov z šeptaných vět se u jednotlivých respondentů pohybovala mezi 52 a 83 % (medián 72 %). U percepčního testu s nahrávkami zpěvu v šepotu mezi 66 a 94 % (medián 83 %). Pro analýzu úspěšnosti v různých kategoriích bylo nutné nejdříve ošetřit pseudoreplikaci. V jednotlivých podskupinách jsem proto vždy podíl správných a všech odpovědí dělila počtem prezentovaných položek, čímž jsem spočítala průměr úspěšností jednotlivých respondentů pro danou skupinu. Ke každému takto vypočtenému průměru jsem pomocí binomického testu v R s použitím balíčku Hmisc (Harrell Jr, 2022) doplnila konfidenční intervaly na hladině významnosti  $\alpha = 0.05$  s Bonferroniho korekcí podle počtu dílčích skupin v každé kategorii.

#### 3.1.1 Úspěšnost v závislosti na zamýšlené melodii

V obou percepčních testech měli respondenti na výběr z tlačítek „stoupnutí“ a „klesnutí“. Obrázek 3.1 ukazuje průměrný podíl správných odpovědí v závislosti na melodii dané položky (stoupavý interval – druhý tón vyšší než první/klesavý interval – druhý tón nižší než první) s konfidenčními intervaly na hladině významnosti  $\alpha = 0.05$  s Bonferroniho korekcí pro  $n = 2$  podle těchto dvou možností odpovědi v testu.

U percepčního testu, který obsahoval slova z šeptaných vět, vyšly správné odpovědi u klesavých intervalů na hladině významnosti  $\alpha = 0.05$  jako statisticky významné. Naopak konfidenční interval pro podíl správných odpovědí u stoupavých intervalů zahrnuje hranici náhody 0.5 a nelze tedy vyloučit, že šlo o pouhé tipování. U percepčního testu zahrnujícího nahrávky zpěvu v šepotu se podíl správných odpovědí u těchto dvou skupin lišil méně a rozpoznání obou typů intervalů vyšlo na hladině významnosti  $\alpha = 0.05$  jako statisticky významné.

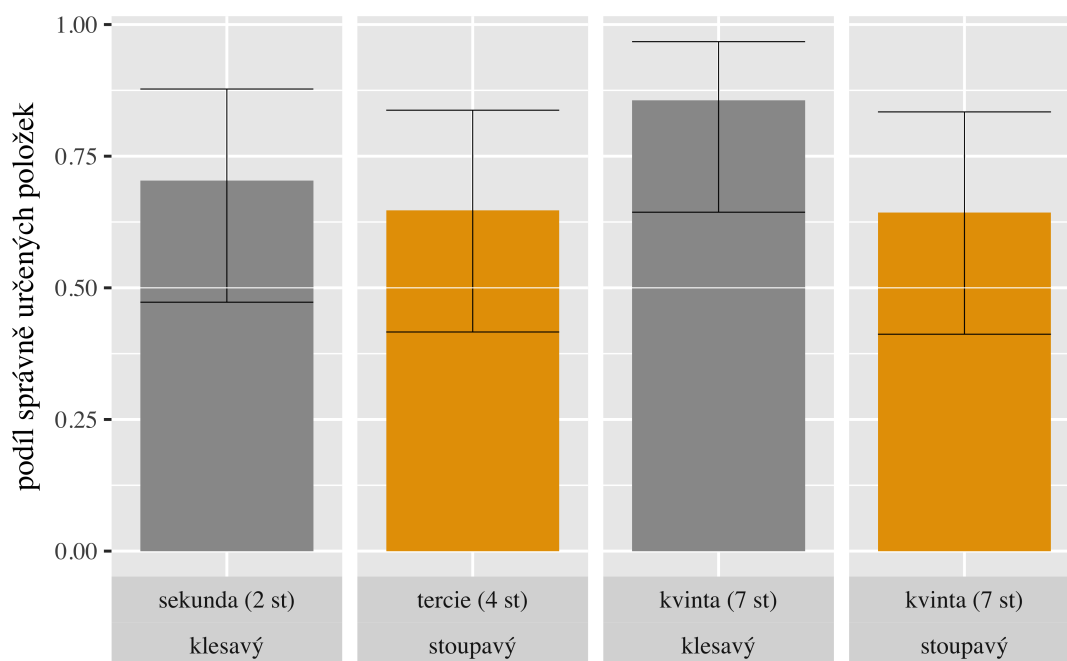


Obrázek 3.1: Průměrný podíl odpovědí, které se shodovaly se záměrem mluvčí, podle směru nahraného intervalu s konfidenčními intervaly na hladině významnosti  $\alpha = 0.05$  s Bonferroniho korekcí pro  $n = 2$  pro každý z percepčních testů. Horizontální předěl značí hranici náhodného tipování.

### 3.1.2 Úspěšnost v závislosti na zamýšleném intervalu

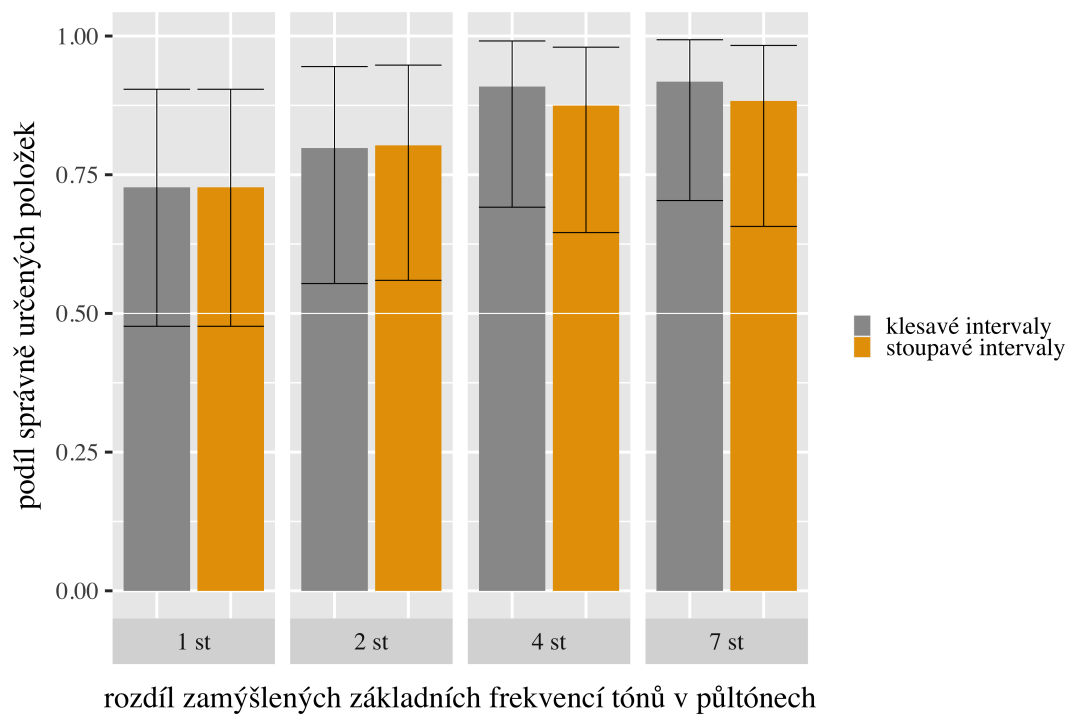
Obrázek 3.2 ukazuje průměrný podíl správných odpovědí v percepčním testu obsahujícím slova z šeptaných vět s konfidenčními intervaly na hladině významnosti  $\alpha = 0.05$  s Bonferroniho korekcí pro  $n = 4$  podle počtu intervalů. Pouze u klesavého intervalu s největším rozdílem v zamýšlené pomyslné základní frekvenci (7 st) se podíl správně určených položek s danou mírou pravděpodobnosti ukazuje jako statisticky významný.

Obrázek 3.3 ukazuje průměrný podíl správných odpovědí v percepčním testu obsahujícím intervaly zpívané v šepotu s konfidenčními intervaly na hladině významnosti  $\alpha = 0.05$  s Bonferroniho korekcí pro  $n = 8$  podle počtu intervalů. U položek s nejmenším rozdílem mezi zamýšlenými tóny (1 st) konfidenční intervaly pro stoupavý i klesavý interval zahrnují hodnotu 0.5 a nelze tedy vyloučit, že odpovědi byly náhodné. Rozpoznání ostatních intervalů (2 st, 4 st, 7 st) ve stoupavé i klesavé podobě bylo s mírou pravděpodobnosti 95 % statisticky významné.

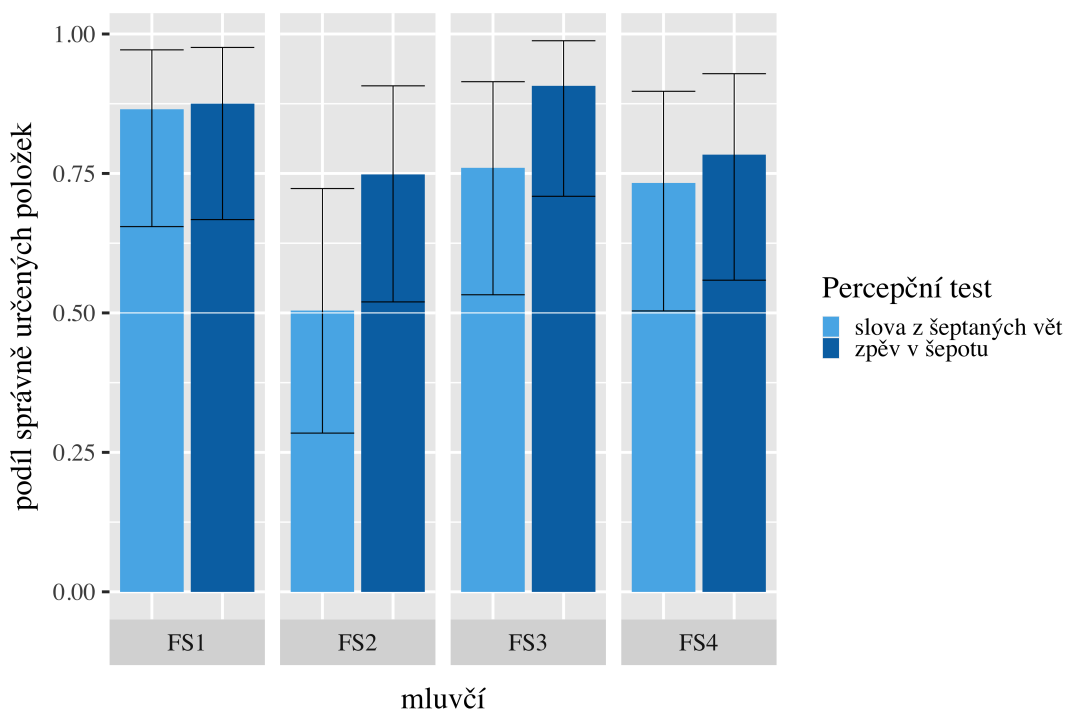


rozdíl zamýšlených základních frekvencí tónů v půltónech a podoba intervalu

Obrázek 3.2: Průměrný podíl odpovědí, které se shodovaly se záměrem mluvčí, v percepčním testu obsahujícím nahrávky šeptaných slov z vět, rozdělené podle směru intervalu a rozdílu v pomyslné  $f_0$  mezi dvěma tóny v rámci položky, doplněné o konfidenční intervaly na hladině významnosti  $\alpha = 0.05$  s Bonferroniho korekcí pro  $n = 4$ . Horizontální předěl značí hranici náhodného tipování.



Obrázek 3.3: Průměrný podíl odpovědí, které se shodovaly se záměrem mluvčí, v percepčním testu obsahujícím nahrávky zpěvu v šepotu, rozdělené podle směru intervalu a rozdílu v pomyslné  $f_0$  mezi dvěma tóny v rámci položky, doplněné o konfidenční intervaly na hladině významnosti  $\alpha = 0.05$  s Bonferroniho korekcí pro  $n = 8$ . Horizontální předěl značí hranici náhodného tipování.



Obrázek 3.4: Průměrný podíl odpovědí v percepčních testech, které se shodovaly se záměrem mluvčích, rozdělené podle jednotlivých mluvčích, doplněné o konfidenční intervaly na hladině významnosti  $\alpha = 0.05$  s Bonferroniho korekcí pro  $n = 4$ . Horizontální předěl značí hranici náhodného tipování.

### 3.1.3 Úspěšnost v závislosti na mluvčích

Obrázek 3.4 ukazuje průměrný podíl správně určených položek v obou percepčních testech v závislosti na mluvčích s konfidenčními intervaly na hladině významnosti  $\alpha = 0.05$  s Bonferroniho korekcí pro  $n = 4$  podle počtu mluvčích. Konfidenční interval průměrné úspěšnosti určení položky u mluvčí FS2 v percepčním testu zahrnujícím slova z šeptaných vět obsahuje hodnotu 0.5 a nelze tedy vyloučit, že odpovědi v tomto případě byly zcela náhodné. Průměrný podíl správně určených položek u této mluvčí v percepčním testu zahrnujícím intervaly zpívané v šepotu a u ostatních mluvčích v obou percepčních testech se už s mírou pravděpodobnosti 95 % ukazují jako statisticky významné. Je však patrné, že mezi jednotlivými mluvčími a percepčními testy jsou mírné rozdíly.

### 3.1.4 Zkušenost s hudbou a rozpoznání melodie v šepotu

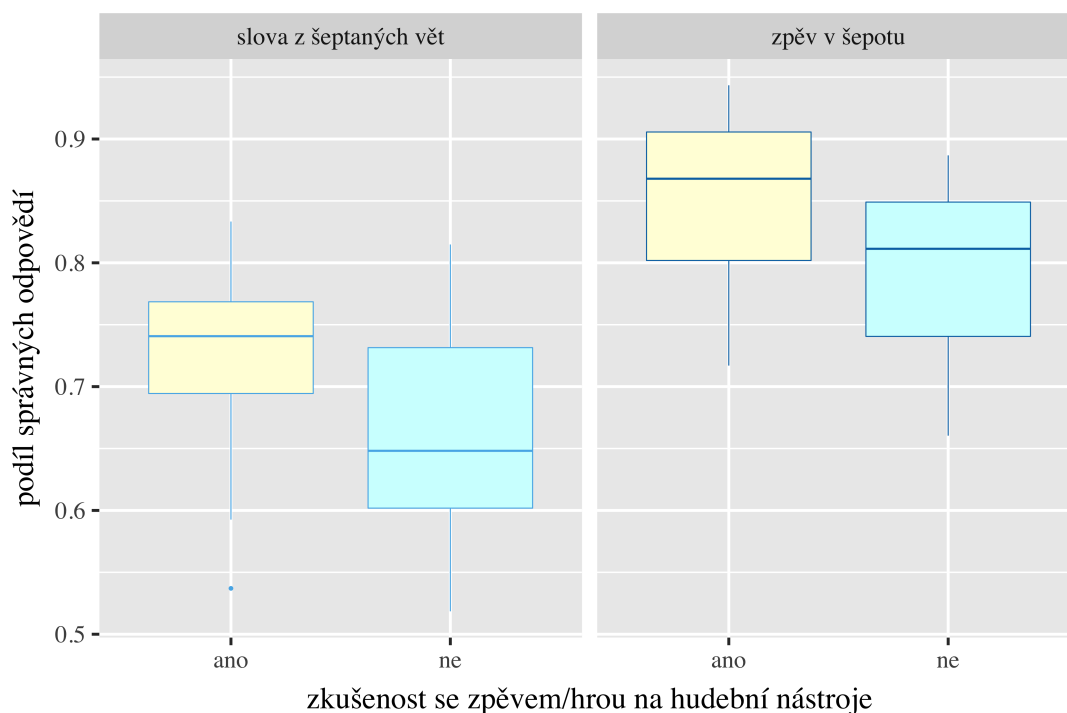
Jednou z doplňujících otázek v úvodu percepčního testu bylo, zda má respondent zkušenosti se zpěvem či hrou na hudební nástroje. Respondentů, kteří na tuto otázku odpověděli „ano“, bylo 23, těch, kteří jako odpověď vybrali „ne“, bylo 10. Rozhodla jsem se proto pro zajímavost otestovat, zda je úspěšnost rozeznání zamýšlené melodie v šepotu ovlivněna touto hudební zkušeností. Po kontrole normálního rozdělení úspěšností v těchto dvou skupinách pomocí Q-Q grafů a Shapiro-Wilksova testu a provedení F-testu jsem pro každý percepční test provedla jednostranný dvouvýběrový t-test nezávislých výběrů se shodností rozptýlů s alternativní hypotézou, že respondenti s hudební zkušeností dosahují vyšší úspěšnosti. Na základě výsledků těchto testů lze u obou percepčních testů na hladině významnosti  $\alpha = 0.05$  zamítnout nulovou hypotézu a přijmout alternativní hypotézu, která říká, že přítomnost hudebních zkušeností má pozitivní vliv na rozeznání zamýšlené melodie v šepotu (pro percepční test se slovy z vět  $p = 0.037$ , pro test s intervaly zpívanými v šepotu  $p = 0.0148$ ). Obrázek 3.5 ukazuje podíl správných odpovědí v percepčních testech podle toho, zda respondenti uvedli, že mají, nebo nemají zkušenosti se zpěvem či hrou na hudební nástroje.

K analýze vlivu hudební zkušenosti na úspěšnost rozeznání zamýšlené melodie v šepotu je však nutné podotknout, že tvrzení respondentů o vlastní hudební zkušenosti v dotazníku nebylo žádným způsobem ověřováno ani nebylo přesněji definováno, co je „zkušeností se zpěvem či hrou na hudební nástroj/e“ míněno. Ověření těchto výsledků by proto vyžadovalo detailnější studii s větším důrazem na hudební pozadí respondentů.

## 3.2 Akustická analýza šepotu

### 3.2.1 F1-F3

První formant nebyl na hladině významnosti  $\alpha = 0.05$  statisticky významně ovlivněn zamýšleným směrem pohybu  $f_0$  a pozicí v intervalu (dále souhrnně na-



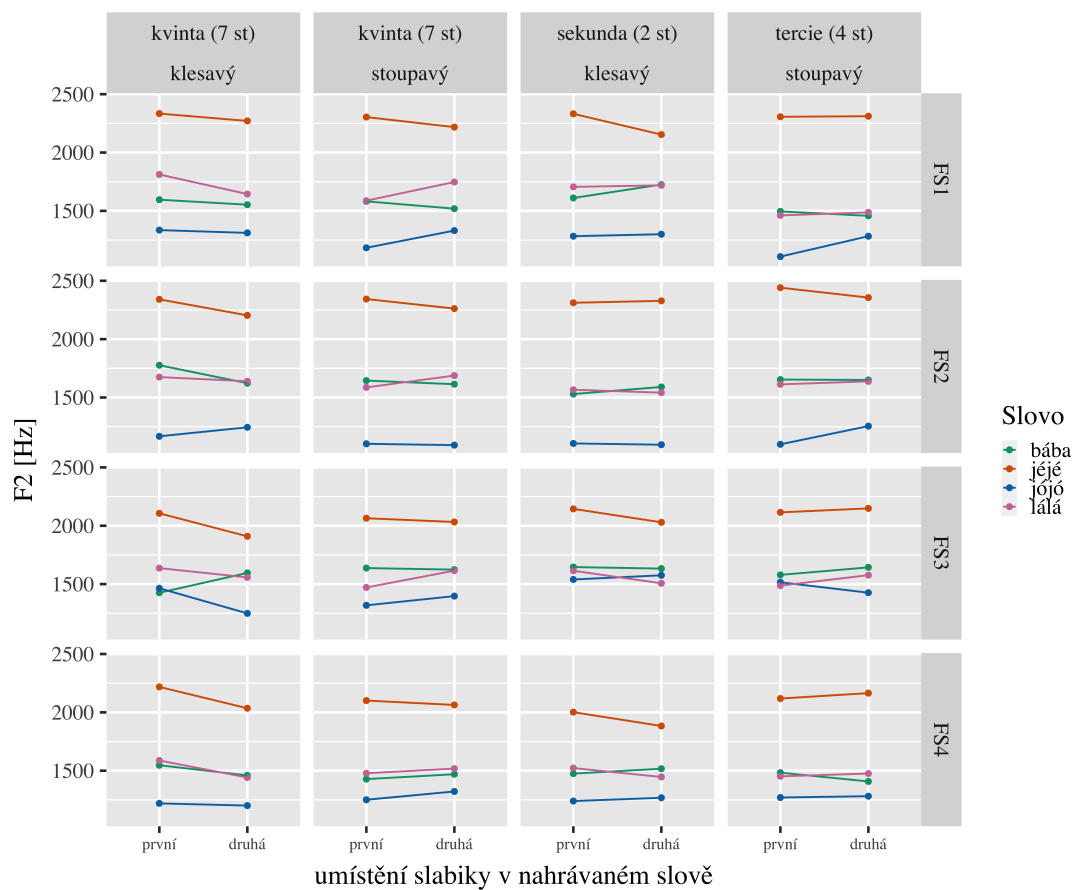
Obrázek 3.5: Podíl správných odpovědí v obou percepčních testech s ohledem na to, zda respondenti v úvodním dotazníku uvedli, že mají, nebo že nemají zkušenosti se zpěvem či hrou na hudební nástroje.

zýváno zamýšlenou melodií) ani u slov z šeptaných vět ( $p = 0.89$ ) ani u zpívaných intervalů ( $p = 0.076$ ). S danou mírou pravděpodobnosti tedy nelze vyloučit, že zamýšlená melodie v šepotu se v prvním formantu neprojeví.

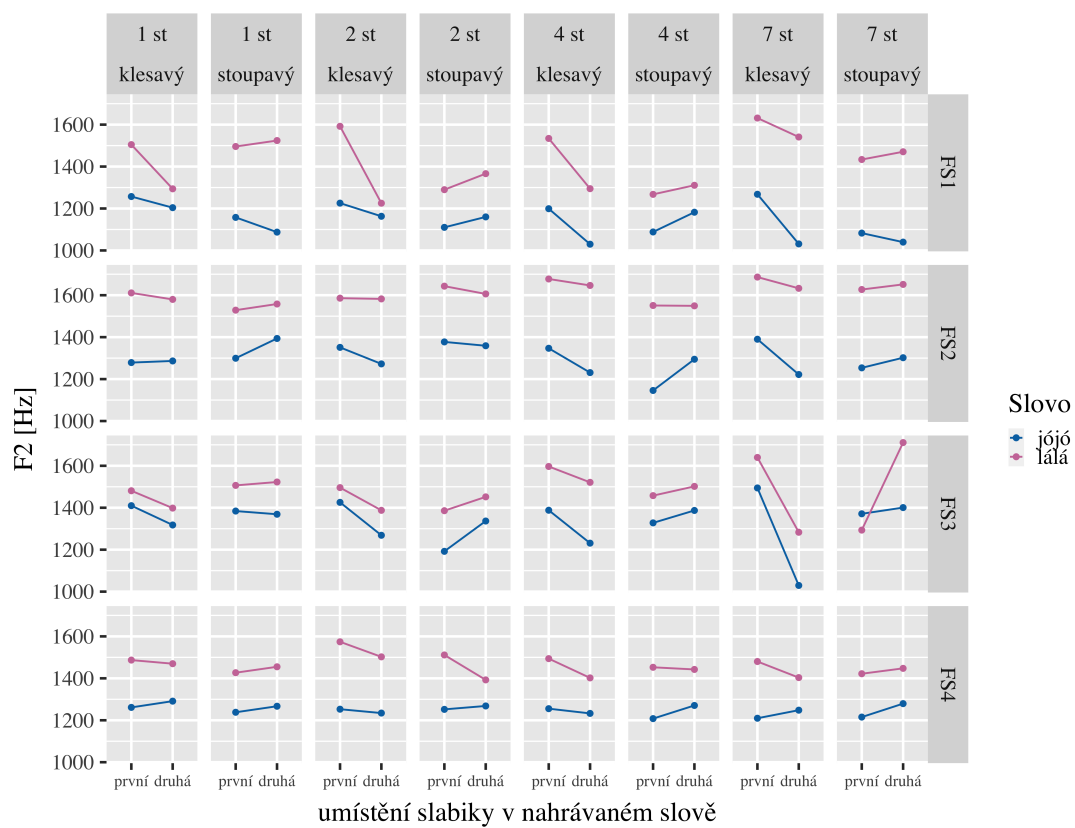
Hodnoty F2 naopak byly na hladině významnosti  $\alpha = 0.05$  statisticky významně ovlivněny zamýšlenou melodií u obou typů nahrávek (pro slova z šeptaných vět  $p = 0.046$ , pro zpěv v šepotu  $p < 0.001$ ). U slov z šeptaných vět i u zpívaných intervalů docházelo ve stoupavých intervalech k nárůstu F2 u druhého tónu, u druhého tónu v klesavých intervalech potom k poklesu. Obrázky 3.6 a 3.7 zobrazují grafy naměřených průměrných hodnot F2 z prostřední třetiny vokálů u slov z šeptaných vět a u zpěvu v šepotu podle jednotlivých intervalů a mluvčích. Hodnoty z první a druhé slabiky každého slova jsou vždy pro lepší orientaci navzájem propojeny.

Třetí formant byl statisticky významně ovlivněn zamýšlenou melodií pouze u zpěvu v šepotu ( $\alpha = 0.05$ ,  $p = 0.008$ ). U stoupavých intervalů docházelo ve druhé slabice k nárůstu frekvence F3, v druhé slabice klesavých intervalů k po-

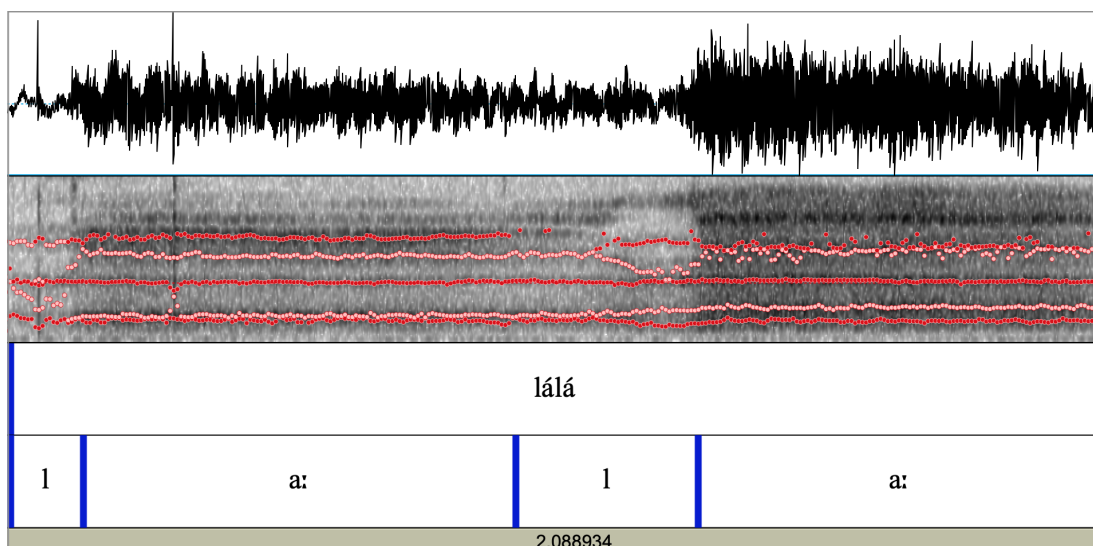




Obrázek 3.6: Průměrné hodnoty F2 z prostřední třetiny vokálů slov z šeptaných vět rozdělené podle podoby intervalu a mluvčích. Hodnoty F2 z vokálů v první a druhé slabice jednoho slova jsou vždy propojeny čarami.



Obrázek 3.7: Průměrné hodnoty F2 z prostřední třetiny vokálů slov zpívaných v šepotu rozdělené podle podoby intervalu a mluvčích. Hodnoty F2 z vokálů v první a druhé slabice jednoho slova jsou vždy propojeny čarami.



Obrázek 3.8: Slovo „lálá“ zpívané v šepotu na stoupavém intervalu s rozdílem 7 st mezi pomyslnou základní frekvencí tónů. Červené tečky značí formanty. Spektrogram zobrazuje frekvence od 0 do 8 kHz, časové rozlišení je 5 ms.

klesu. U slov z šeptaných vět nelze na hladině významnosti  $\alpha = 0.05$  vyloučit, že zamýšlená melodie na F3 nemá vliv ( $p = 0.059$ ).

Obrázek 3.8 ukazuje spektrogram jedné z nahrávek zpívaných intervalů, která byla v percepčním testu správně určena 100 % respondentů (7 st stoupavý, mluvčí FS3, slovo „lálá“), se zvýrazněnými formanty. Mezi F1 a F3 není okem pozorovatelný rozdíl napříč oběma vokály. F3 ovšem vyšel jako statisticky významný parametr a skutečně je podle naměřených údajů u druhého tónu o 2 % vyšší (nárůst o 62 Hz). Linie F2 je však u vyššího z tónů viditelně posunutá směrem nahoru, v tomto případě se jedná o nárůst o 32 % (416 Hz).

### 3.2.2 Poměry jednotlivých formantů

U šeptaných slov z vět i zpěvu v šepotu jsem zkoumala vliv zamýšlené melodie na poměr všech kombinací prvního až třetího formantu. U slov z šeptaných vět na hladině významnosti  $\alpha = 0.05$  nelze zamítnout, že žádný z těchto poměrů formantů zamýšlenou melodií ovlivněn není (pro F2:F1  $p = 0.27$ , pro F3:F2  $p = 0.41$  a pro F3:F1  $p = 0.43$ ). U zpěvu v šepotu byl však poměr F2 s oběma dalšími formanty s mírou pravděpodobnosti 95 % zamýšlenou melodií statisticky

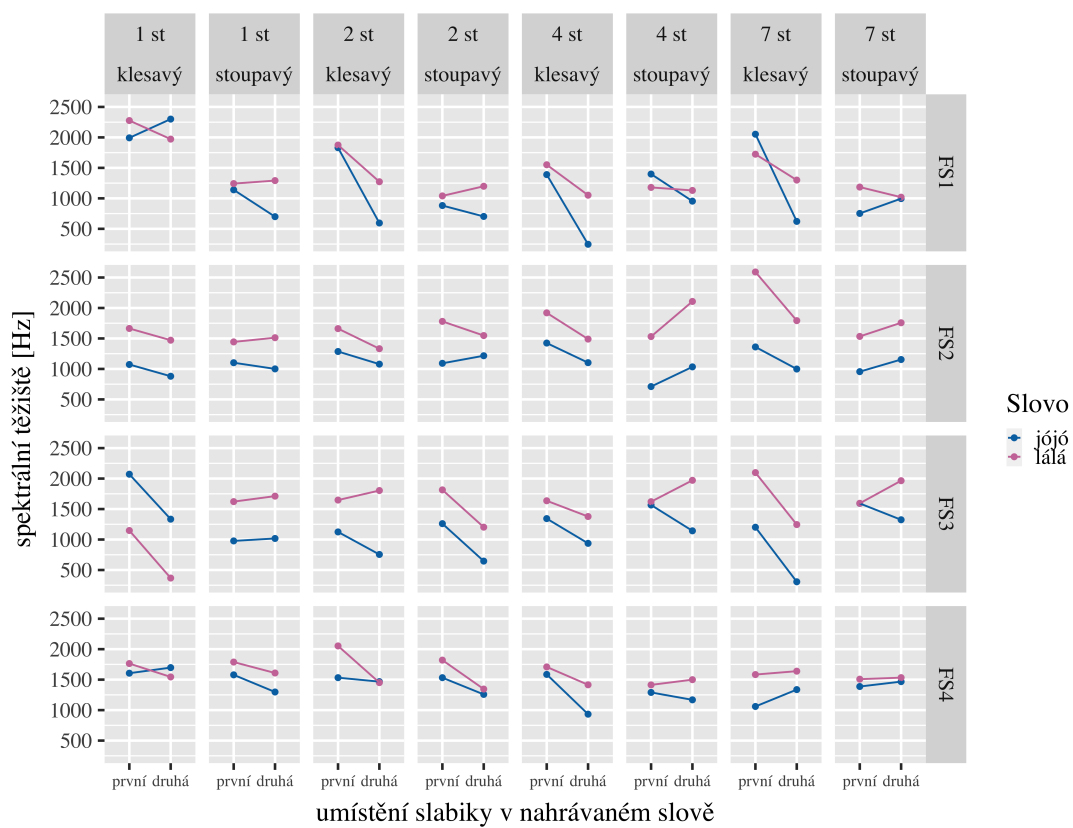
významně ovlivněný (pro F2:F1  $p < 0.001$ , pro F3:F2  $p = 0.012$ ). Poměr F2:F1 se ve druhé slabice u klesavých intervalů zmenšuje a u stoupavých zvětšuje, což značí větší rozdíl mezi F2 a F1 u vyššího zamýšleného tónu. Naopak poměr F3:F2 se ve druhé slabice u klesavých intervalů zvětšuje a u stoupavých zmenšuje, což značí menší rozdíl mezi F3 a F2 u vyššího zamýšleného tónu. Tyto dvě vlastnosti odpovídají pohybu F2 směrem nahoru u vyššího zamýšleného tónu.

Stejně jako u slov z šeptaných vět nelze ani u zpěvu v šepotu na hladině významnosti  $\alpha = 0.05$  vyloučit, že poměr frekvencí třetího ku prvnímu formantu zamýšlenou melodií ovlivněn není ( $p = 0.64$ ).

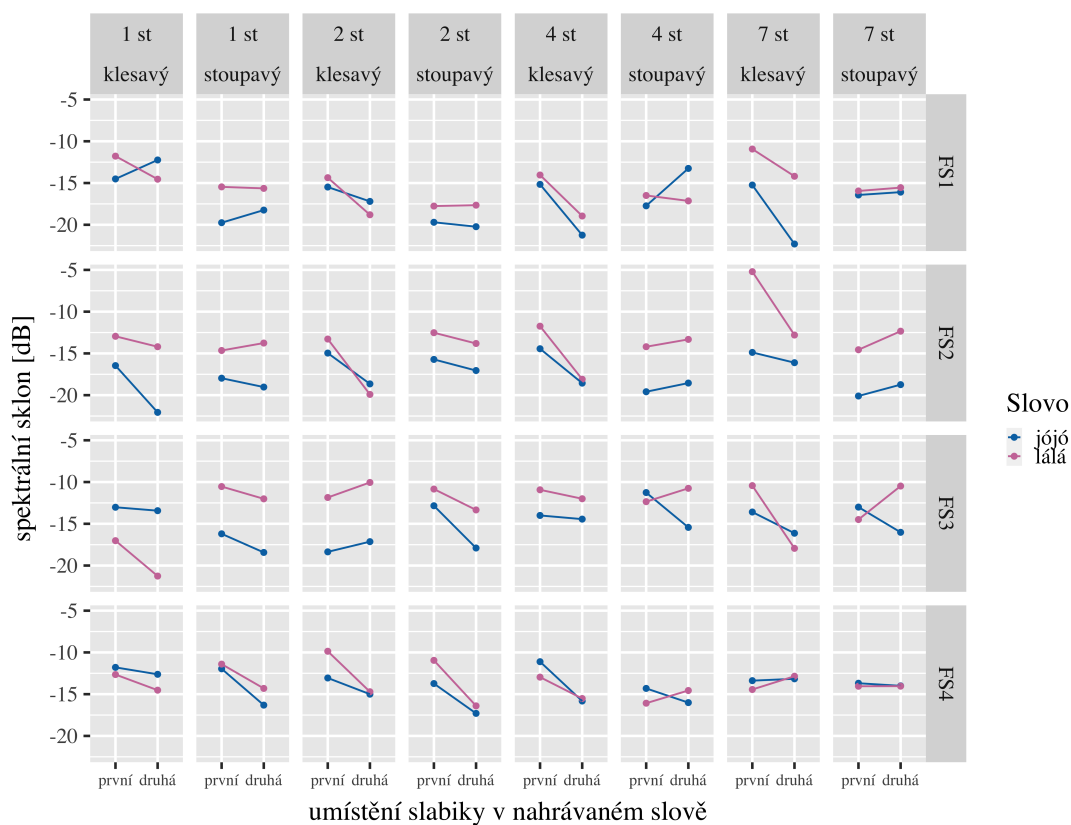
### 3.2.3 Spektrální těžiště

U intervalů zpívaných v šepotu bylo spektrální těžiště z prostřední třetiny vokálů na hladině významnosti  $\alpha = 0.05$  statisticky významně ovlivněno zamýšlenou melodií ( $p = 0.0069$ ). Pokles spektrálního těžiště vokálů v druhé slabice klesavých intervalů byl výraznější než nárůst spektrálního těžiště vokálů v druhé slabice stoupavých intervalů. Obrázek 3.9 ukazuje graf naměřených hodnot spektrálního těžiště v Hz v přehledu podle podoby intervalu a mluvčích. Hodnoty z vokálů první a druhé slabiky jednoho slova jsou vždy vizuálně propojeny.

Spektrální těžiště spektra prostřední třetiny vokálů slov z šeptaných vět naměřené v Praatu z celkového signálu se na hladině významnosti  $\alpha = 0.05$  neprokázalo jako parametr statisticky významně ovlivněný zamýšlenou melodií ( $p = 0.056$ ) a s danou mírou pravděpodobnosti tedy nelze vyloučit, že jí ovlivněno není. Spektrální těžiště stejného signálu upraveného filtrem pásmová zadrž mezi 1 a 6.3 kHz, který odstranil frekvenční pásmo, ve kterém se vyskytovala hlavní formantová struktura vokálů viditelná ve spektrogramu, se však už s mírou pravděpodobnosti 95 % ukázal jako statisticky významně ovlivněný zamýšlenou melodií ( $p = 0.022$ ). Praktická významnost a možné zobecnění tohoto výsledku by však bylo nutné ověřit akustickou analýzou jiné sady dat a percepčním testem s filtrovanými nahrávkami.



Obrázek 3.9: Spektrální těžiště prostřední třetiny vokálů z intervalů zpívaných v šepotu rozdělené podle podoby intervalu a mluvčích. Hodnoty spektrálního těžiště vokálů z první a druhé slabiky jednoho slova jsou vždy propojeny čarami.



Obrázek 3.10: Spektrální sklon v prostřední třetině vokálů z intervalů zpívaných v šepotu rozdělené podle podoby intervalu a mluvčích. Hodnoty z vokálů z první a druhé slabiky jednoho slova jsou vždy propojeny čarami.

### 3.2.4 Spektrální sklon

Spektrální sklon v prostřední třetině vokálů byl s mírou pravděpodobnosti 95 % statisticky významně ovlivněn zamýšlenou melodií pouze u zpěvu v šepotu ( $p = 0.0069$ ). U stoupavých i klesavých intervalů docházelo ve druhé slabice slova k poklesu, pokles v klesavých intervalech byl však oproti poklesu ve stoupavých intervalech čtyřnásobný (všechny hodnoty byly v záporných číslech, pokles v tomto případě tedy znamená větší negativní zápornou hodnotu). Obrázek 3.10 ukazuje graf naměřeného spektrálního sklonu v dB v přehledu podle podoby intervalu a mluvčích. Hodnoty z vokálů první a druhé slabiky jednoho slova jsou vždy navzájem propojeny čarou.

U slov z šeptaných vět nelze na hladině významnosti  $\alpha = 0.05$  vyloučit, že zamýšlená melodie neměla na spektrální sklon vliv ( $p = 0.24$ ).

### 3.3 Akustické koreláty a percepční test

U nahrávek slov z šeptaných vět se na hladině významnosti  $\alpha = 0.05$  jako akustické parametry statisticky významně ovlivněné zamýšlenou melodií ukázaly F2 a spektrální těžiště signálu filtrovaného filtrem pásmová zádrž mezi 1 a 6.3 kHz, u intervalů zpívaných v šepotu potom druhý a třetí formant, poměr F2:F1 a F3:F2, spektrální těžiště a spektrální sklon. Naměřené hodnoty těchto parametrů v jednotlivých vokálech jsem s použitím programu R spárovala podle nahrávaného slova, jeho melodie a mluvčí, čímž jsem pro každé slovo, které bylo nahráno jako jeden celek, získala hodnoty v jeho první a druhé slabice. Následně jsem pro každou jednotlivou nahrávku v případě formantů a spektrálního těžiště vypočetla poměr těchto parametrů postupem, kdy jsem vždy dělila hodnotu získanou z vyššího tónu hodnotou z nižšího (tedy u stoupavého intervalu jsem dělila hodnotu z druhé slabiky hodnotou z první a u klesavého naopak hodnotu z první slabiky hodnotou z druhé). V případě spektrálního sklonu a poměrů formantů, které už samy o sobě jsou poměry, jsem vždy vypočítala rozdíl těchto parametrů mezi dvěma vokály jednoho slova. U položek z percepčního testu, který obsahoval slova z šeptaných vět, jsem vyřadila položky namluvené mluvčí FS2, u které konfidenční interval podílu správně určených položek zahrnoval hranici náhodného tipování.

Přesto, že u nejhůře rozpoznávaných položek (správně určených pod 50 % respondentů) byly poměry či rozdíly naměřených hodnot statisticky významných parametrů menší, u položek s nejlepší rozpoznatelností (správně určené 95-100 % respondentů) byly zastoupeny velké i malé poměry mezi hodnotami. Korelační grafy u žádného samostatného poměru hodnot akustických parametrů neukázaly jednoduchý lineární vztah s úspěšností určení dané položky napříč respondenty (proto jsem je do této práce ani nezahrnula). Lze se tedy dle očekávání domnívat, že žádný z parametrů v této sadě nahrávek není sám o sobě jediným vypovídajícím faktorem z hlediska rozpoznatelnosti zamýšlené melodie, jejíž určení se zdá být velmi komplexní záležitostí.

### 3.4 Chování parametrů v modální fonaci

Stejně jako u šepotu se u modální fonace na hladině významnosti  $\alpha = 0.05$  potvrdil statisticky významný vliv melodie na F2 (pro řeč  $p = 0.005$ , pro zpěv  $p = 0.0076$ ). Dále byl s mírou pravděpodobnosti 95 % ovlivněn i první formant (pro řeč i pro zpěv  $p < 0.001$ ). Z poměrů formantů byl v řeči i ve zpěvu melodií statisticky významně ovlivněn poměr F3:F2 u zpěvu ( $\alpha = 0.05$ ,  $p = 0.001$ ) a u obou typů projevu poměr F3:F1 ( $\alpha = 0.05$ , pro řeč  $p = 0.002$ , pro zpěv  $p < 0.001$ ). Dalším parametrem na hladině významnosti  $\alpha = 0.05$  statisticky významně ovlivněným melodií v řeči i ve zpěvu bylo spektrální těžiště (u obou  $p < 0.001$ ). U řeči potom také spektrální sklon ( $\alpha = 0.05$ ,  $p < 0.001$ ). Spektrální sklon ve zpívaných vokálech se s mírou pravděpodobnosti 95 % jako statisticky významně ovlivněný melodií neprojevil ( $p = 0.084$ ).

Na nahrávky modální fonace byl však při jejich pořizování kladen menší důraz a sloužily především k tomu, aby se mluvčí seznámily s charakterem zadávaných úkolů. Naměřené hodnoty a jejich statistickou významnost je proto nutné brát jen orientačně.



## 4. Diskuse

V posledním desetiletí nastal velmi rychlý vývoj různých domácích technologií a mobilních aplikací, které pracují s rozpoznáváním a syntézou řeči a velmi často jsou primárně ovládány hlasem. Některé z nich byly na základě výzkumu počítačové analýzy a syntézy šepotu nedávno obohaceny o režim šepotu (whisper mode), kdy člověk komunikuje se svým přístrojem v šepotu a ten mu šepem odpovídá (Raeesy, 2018). Jelikož suprasegmentální vlastnosti řeči jsou zásadní pro přirozenost projevu, zabývá se související výzkum mimo jiné i melodií v šepotu. Výzkum zabývající se touto problematikou se sice postupně rozšiřuje, je však především cílen na praktickou využitelnost získaných poznatků k vylepšení takového přístroje, a tak kromě faktu, že nejčastěji zkoumá v počítačové oblasti nejrozšířenější angličtinu, často řeší jiné výzkumné otázky, než jsou běžné ve fonetickém výzkumu, jako například jak spolehlivě v reálném čase automaticky odlišit šepot od jiného typu fonace (Raeesy a kol., 2018) nebo jak syntetizovat šepot na základě nahrávek modální fonace (Cotescu a kol., 2020).

Novodobý fonetický výzkum melodie v šepotu se uskutečňuje především v prostředí evropských jazyků, jako je například holandština (Heeren, 2015; Heeren a van Heuven, 2014) nebo francouzština (Heeren a Lorenzi, 2014). Není mi známo, že by se vyskytla obdobná studie zabývající se melodií v šepotu v českém prostředí, při práci na této bakalářské práci jsem proto vycházela ze zdrojů zpracovávajících melodie v šepotu v jiných jazycích s ohledem na informace o intonaci v modální fonaci v češtině. Zpracování akustické analýzy vokálů ve slovech z šepotaných vět a intervalů zpívaných v šepotu v rámci této práce přinesly zajímavé výsledky, které se v některých aspektech podobají dosavadnímu výzkumu, v jiných se od něj naopak liší.

## 4.1 Akustické koreláty melodie v šepotu

Během zpracování výsledků jsem pomocí lineárních smíšených modelů analyzovala vliv zamýšlené melodie v šepotu a reálné melodie v modální fonaci na různé akustické parametry. Komentář k výsledkům akustické analýzy šepotu podávám níže, vyvarovala jsem se však přímému srovnání výsledků analýzy hodnot naměřených v šepotu s analýzou hodnot naměřených v řeči v modální fonaci a ve zpěvu. Verze nahrávaných položek v modální fonaci byla určena především k tomu, aby se mluvčí seznámily s postupem nahrávání materiálu a za účelem možného pozdějšího srovnání  $f_0$  vokálů v modální fonaci s předlohou k nahrávání. Bloky modální fonace a šepotu se v rámci nahrávání jedné mluvčí střídaly, byly tak nahrány se stejným nastavením vstupního zesilovače. Některé mluvčí se však zvláště během nahrávání zpěvu občas pohnuly blíže k mikrofonu, a tak u některých nahrávek nedopatřením došlo ke zkreslení nahraného zvuku (clipping). Výsledky analýzy modální fonace je proto třeba brát pouze jako orientační. Vzhledem k tomu, že analýza modální fonace nebyla hlavní náplní této práce, věnuji se v následující diskusi výsledků téměř výhradně akustickým korelátům melodie v šepotu.

### 4.1.1 Formanty a poměry formantů

Analýza vlivu zamýšlené melodie v šepotu na formanty prokázala u slov z šepotaných vět vliv pouze u formantu druhého ( $\alpha = 0.05$ ,  $p = 0.046$ ), u zpěvu v šepotu se potom prokázal vliv u F2 ( $\alpha = 0.05$ ,  $p < 0.001$ ), F3 ( $\alpha = 0.05$ ,  $p = 0.008$ ) a také poměrů F2:F1 ( $\alpha = 0.05$ ,  $p < 0.001$ ) a F3:F2 ( $\alpha = 0.05$ ,  $p = 0.012$ ), kde vliv zamýšlené melodie na tyto poměry formantů poukazuje na znatelnější pohyb u druhého než u prvního a třetího formantu. Pohyb F2 v závislosti na zamýšlené melodii se shoduje se závěry I. B. Thomase (1969), který na základě percepčního experimentu určil frekvenci F2 jako absolutní vnímanou výšku vokálů v šepotu. Z Thomasových závěrů lze usuzovat, že pokud se F2 mění, vnímaná výška vokálu se bude také měnit, a tím by druhý formant mohl být v šepotu vnímán jako

náhrada pomyslné  $f_0$ .

Pohyb F2 a jeho vztah s okolními formanty byl u zpěvu v šepotu znatelnější než u šeptané řeči. Možné vysvětlení můžeme hledat například v charakteristikách běžného zpěvu, kde sledujeme mírně odlišné záměry než u řeči v modální fonaci. Zatímco intonací v řeči komunikujeme význam a nejde o konkrétní absolutní hodnotu  $f_0$ , ve zpěvu je přesná melodie žádoucí a odlišnost kolem jednoho půltónu od tónu, který máme v úmyslu realizovat, už je vnímána jako melodie odlišná. V řeči máme možná také jiné preference z hlediska srozumitelnosti sdělení, a tak více než ve zpěvu klademe důraz na zachování kvality, a tedy i rozeznatelnosti, konkrétního vokálu. Ve zpěvu je naopak srozumitelnost zpívaných vokálů upozaděna na úkor přesnosti melodie. Tato úvaha tedy vede k domněnce, že F2 je součástí vnímané melodické kontury v šepotu, v řeči však nemůže mít takovou variabilitu, která by ovlivnila vnímanou kvalitu daného vokálu. Je také třeba podotknout, že vokály ve zpívaných intervalech měly výrazně delší trvání než vokály ve slovech z šeptaných vět, a mluvčí tak měly větší prostor na dosažení artikulačního cíle.

Význam F2 jako korelátu melodie v šepotu potvrzují i závěry novějších studií (výjimku tvoří jedna ze studií melodie v šeptané holandštině, kde byl pohyb F2 v závislosti na zamýšlené melodii pozorovatelný, neukázal se však jako statisticky významný, viz Heeren a van Heuven (2014)). Percepční testy se syntetizovanými vokály ukázaly, že melodie je vnímána jako stoupavá, pokud se F2 pohybuje směrem nahoru. Tato vlastnost se umocnila, když se zároveň směrem nahoru pohyboval i F1 (Higashikawa a Minifie, 1999). V souladu s tímto poznatkem byl v této práci F1 u řeči i zpěvu melodií statisticky významně ovlivněný v modální fonaci ( $\alpha = 0.05$ , pro řeč i zpěv  $p < 0.001$ ). Další výzkum šeptané holandštiny prokázal statistickou významnost vlivu zamýšlené melodie v šepotu na první až třetí formant (Heeren, 2015). F1 se ovšem v rámci této práce jako významný korelát neukázal v případě slov z šeptaných vět ( $\alpha = 0.05$ ,  $p = 0.89$ ) ani zpěvu v šepotu ( $\alpha = 0.05$ ,  $p = 0.076$ ).

Percepční testy s filtrovanými nahrávkami šeptané francouzštiny potom do-

kázaly, že mezi formanty nelze nalézt korelát, jehož nepřítomnost by znemožnila rozpoznání zamýšlené melodie v šepotu. Výsledky percepčního testu ukázaly, že akustické parametry zásadní pro rozpoznání melodie v šepotu se nejspíše nachází ve frekvencích nad 1500 Hz. Domněnka, že hlavním korelátem melodie je druhý formant, se však nepotvrdila. V kontrolním experimentu pak při odstranění frekvenčního pásma zahrnujícího F2 dosahovali respondenti stejných výsledků jako při odstranění pásma, které zahrnovalo F3 (Heeren a Lorenzi, 2014).

### 4.1.2 Spektrální těžiště

Meyer-Epppler (1957) uvedl jako korelát melodie v šepotu u některých vokálů vysokofrekvenční spektrální šum. Dalo by se očekávat, že pokud je míra přítomnosti takového šumu akustickým parametrem odrážejícím šeptanou intonaci, projeví se tato skutečnost v posunu spektrálního těžiště u vyšších tónů směrem nahoru. Vliv zamýšlené melodie na spektrální těžiště zkoumají studie využívající novější počítačové technologie umožňující jeho výpočet (např. Praat). Spektrální těžiště se neukázalo jako významný korelát melodie v jedné studii šeptané holandštiny (Heeren a van Heuven, 2014), pozdější studie naopak statisticky významný vliv zamýšlené melodie na spektrální těžiště potvrdila (Heeren, 2015).

V případě této bakalářské práce bylo spektrální těžiště zamýšlenou melodií ovlivněno u vokálů z nahrávek zpěvu v šepotu ( $\alpha = 0.05$ ,  $p = 0.002$ ). U vokálů ze slov z šeptaných vět byl tento vliv na hranici statistické významnosti ( $\alpha = 0.05$ ,  $p = 0.053$ ). Nelze tedy vyloučit, že s větším množstvím nahrávek či jinou sadou dat by bylo možné tento vztah v šeptané češtině prokázat.

U slov z šeptaných vět bylo statisticky významně ovlivněno spektrální těžiště signálu po odstranění pásma od 1 do 6.3 kHz, ve kterém se nacházely hlavní rozlišitelné formantové struktury, filtrem pásmová zádrž ( $\alpha = 0.05$ ,  $p = 0.022$ ). To by mohlo naznačovat již zmíněnou přítomnost akustických vodítek melodie v šepotu v pásmu vysokých frekvencí. Bez dalšího ověření této vlastnosti akustickou analýzou či percepčním testem však nelze s jistotou říci, že by tento výsledek byl

vypovídající o akustických projevech melodie v šepotu.

### 4.1.3 Spektrální sklon

Spektrální sklon je používán jako ukazatel mluvního úsilí a v některých jazycích je korelátem přízvučné slabiky (Skarnitzl a kol., 2016, str. 144-145). Percepční experimenty se signály, které se lišily pouze spektrálním sklonem, ukázaly, že spektrální sklon může ovlivnit vnímanou výšku signálu (Kuang a Liberman, 2015). Zároveň se potvrzuje, že by spektrální sklon mohl být jedním z akustických parametrů ovlivněných zamýšlenou melodií v šepotu (Heeren, 2015). Toto zjištění je v souladu s dřívějším předpokladem, že pokud by produkce vyšších tónů znamenala větší úsilí než produkce tónů nižších, dá se u vyšších tónů očekávat menší (méně negativní) spektrální sklon (Heeren a van Heuven, 2014).

Výsledky akustické analýzy v rámci této bakalářské práce ukázaly, že vliv zamýšlené melodie na spektrální sklon v šeptané řeči se neblíží statistické významnosti ( $\alpha = 0.05$ ,  $p = 0.24$ ), naopak pro zpěv se spektrální sklon jeví jako parametr zamýšlenou melodií statisticky významně ovlivněný ( $\alpha = 0.05$ ,  $p = 0.0069$ ). Možné vysvětlení by bylo v charakteru zpěvu, který se na rozdíl od řeči neuskutečňuje v přirozené výšce (poloze) hlasu, ale spíš výše. Při opakování mluvených vět měly mluvčí napodobit intonaci, kterou slyšely, je však pravděpodobné, že produkovaly zadanou intonační konturu ve své přirozené poloze hlasu. Naopak u zpěvu produkovaly přesný tón dle zadání, od kterého se v modální fonaci odchýlily průměrně jen o 2 Hz. Cílová slova v mluvených předlohách se také pohybovala v nižších základních frekvencích než předlohy pro zpěv ( $f_0$  u cílových slov ve větách, nahraných v mojí přirozené poloze hlasu, mezi 175 a 262 Hz;  $f_0$  tónů v předloze pro zpěv mezi 264 a 396 Hz). Dá se tedy předpokládat, že u zpěvu, a tedy i zpěvu v šepotu, bude úsilí v produkci vyšších tónů znatelněji odlišné od tónů nižších. Ve zpěvu v šepotu tak může zamýšlená melodie více ovlivňovat mluvní úsilí u jednotlivých tónů, což se v tomto případě ve spektrálním sklonu projevilo.

## 4.2 Zpracování percepčního testu

### 4.2.1 Výsledky percepčního testu

Dle očekávání byly intervaly s větším rozdílem v pomyslné  $f_0$  mezi dvěma tóny rozeznávány lépe než intervaly s rozdíly minimálními. V obou percepčních testech však zároveň byly lépe rozpoznávány klesavé intervaly než intervaly stoupavé. V percepčním experimentu s minimálními páry šeptaných holandských dvouslabičných slov odlišených přízvukem realizovanými jako oznamovací věty a otázky zjišťovací se projevila podobná tendence, kdy byly položky, kde se umístění vyššího z tónů shodovalo s umístěním přízvuku v daném slově, rozpoznávány lépe než slova, kde byl přízvuk na slabice s nižší pomyslnou  $f_0$ . V běžné mluvené češtině není příliš prominentní důrazový přízvuk (Skarnitzl a kol., 2016, str. 146-147), je však pravděpodobné, že jsem při nahrávání předloh, ovlivněna vědomím, které slovo je v dané větě zásadní, kladla větší důraz na slova realizovaná na cílovém intervalu a realizovala tak důrazový přízvuk výraznější, než by se vyskytl u daného slova ve stejné větě v běžné řeči. Mluvní styl v předlohách byl také spíše pečlivý, zatímco v každodenní řeči by byl častější styl neutrální. Toto zdůraznění mohlo v kombinaci s vyšším tónem přispět k identifikovatelnosti klesavých intervalů.

Rozložení výsledků by bylo možné samozřejmě také vysvětlit preferencí respondentů pro určitý interval. Pokud by například obecnou tendencí bylo chování, kdy v případě nejistoty respondent vždy automaticky položku klasifikuje jako klesavý interval, nepoměrně by se tím zvýšil počet správných odpovědí v této kategorii. Takovémuto zkreslení výsledků jsem se snažila při tvorbě percepčních testů v co největší míře předejít (viz kap. Metoda), bez opakovaných experimentů zkoumajících rozpoznatelnost melodie v šeptané češtině však nelze posoudit efektivnost mých opatření.

Srovnání úspěšností v percepčním testu mezi respondenty se zkušenostmi se zpěvem či hrou na hudební nástroje a respondenty bez hudebních zkušeností je v tomto podání spíše zajímavostí než plnohodnotným výzkumem. Jak zmiňuji

v kapitole Výsledky, zařazení respondentů do těchto kategorií jsem provedla na základě jedné nepříliš specificky definované otázky a nebrala ohled na možnou variabilitu projevů těchto zkušeností u jednotlivých respondentů. Zastoupení jednotlivých skupin také bylo nerovnoměrné, zatímco ve skupině s hudební zkušeností bylo 23 osob, ve skupině bez hudební zkušenosti jich bylo pouze 10. Postup k získání reálně zobecnitelných výsledků by vyžadoval přesnější definice jednotlivých skupin, případně pokročilejší statistické metody.

## 4.2.2 Percepční test v online prostředí

I přes zdánlivý ústup koronavirové pandemie jsem se rozhodla uskutečnit percepční testy zkoumající rozpoznatelnost melodie v šepotu online pomocí nástroje PsyToolkit, který je svým původem určen k přípravě psychologických experimentů (ne vždy nepodobných experimentům fonetickým) a to nejen z důvodu přetrvávajících karantén, ale i pro usnadnění postupu respondentům (kteří tak nemuseli kvůli experimentu cestovat ani například stahovat Praat, pro tvorbu fonetických percepčních testů nejrozšířenější nástroj) a tedy maximalizace počtu možných respondentů v čase, po který experiment probíhal.

Experimenty v tomto nástroji jsou snadno programovatelné a umožňují nespočet téměř libovolných konfigurací. Díky online prostředí je také možnost takovýto percepční test zadávat na dálku různým skupinám respondentů či ho sdílet přes internet s širokou veřejností, což má na jednu stranu svá úskalí v nekontrolovatelnosti účastníků výzkumu, zároveň by se však takto získaná skupina respondentů mohla více blížit náhodnému výběru z populace (např. se znalostí češtiny), samozřejmě pokud to je soubor respondentů, který hledáme.

Výroba percepčního testu v nástroji PsyToolkit vyžaduje si osvojit systém programování experimentů a dotazníků a jejich vzájemného propojení, zároveň je třeba zvláště připravit všechny prvky, které se v rámci testu vyskytnou (př. úvodní strana s instrukcemi, tlačítka). Tento postup by někomu mohl připadat zbytečný či zdouhavý, umožňuje však velmi vysokou variabilitu a možnosti úpravy dle

vlastních představ. V percepčních testech k této bakalářské práci jsem díky vlastní tvorbě prvků mohla snadno přidat návodné ilustrace na tlačítka s odpověďmi. Pokud by se využití tohoto nástroje pro fonetické účely rozšířilo, věřím, že by jistě snadno vznikl soubor sdílených šablon a předpřipravených prvků, které by výrobu percepčního testu usnadnily.

### 4.2.3 Přesah možného výzkumu získaných dat

Kromě odpovědí z percepčních testů lze kromě základních demografických údajů z úvodního dotazníku také velmi snadno získat u každého respondenta informace o rozmístění tlačítek v percepčním testu, pořadí jednotlivých položek (v případě obou percepčních testů v rámci této práce bylo rozmístění tlačítek i pořadí položek vždy náhodně generováno), celkový čas strávený u experimentu, reakční dobu u jednotlivých položek, nebo například i to, zda (u testu s možností několika opakování případně kolikrát) si respondent každou položku přehrál znovu.

Na základě těchto informací lze zkoumat jednak úspěšnost v percepčním testu v různých skupinách (např. muži/ženy či rozdělení podle hudební zkušenosti, viz výše), zároveň je ale možné z nich vyvodit závěry o zkreslení výsledků preferencí určitých odpovědí (angl. bias). Respondenti mohli v případě nejasnosti mít tendenci vybrat vždy odpověď „stoupnutí“ či naopak vždy odpověď „klesnutí“. Mohli také mít preferenci pro jedno ze dvou nabízených tlačítek (levé/pravé, horní/spodní). Nehledě na správnou odpověď u dané položky tak můžeme odhalit obecné preference, které se projeví ve vyplňování percepčního testu, a tyto preference srovnat například napříč různými skupinami.

## 4.3 Další možnosti směřování výzkumu

Základním prvkem vnímané intonace v modální fonaci je průběh  $f_0$  (Skarnitzl a kol., 2016, str. 125). Jsou ale i studie, které ukazují význam konsonantů,



především neznělých frikativ, ve vnímání intonace, jejich akustické chování je označováno jako intonace segmentální (Niebuhr, 2009). V souvislosti s těmito vlastnostmi se začínají objevovat i výzkumy intonace v šepotu (např. studie segmentální intonace v polštině, viz Žygis a kol. (2017)), možnosti experimentů v této oblasti však zdaleka nejsou vyčerpány.

Projev konkrétního mluvčího ovlivňuje, jak vnímáme kvalitu vokálů, které produkuje (Ladefoged a Broadbent, 1957). Zároveň i intonaci vnímáme na základě vztahů akustických parametrů v rámci projevu mluvčího. Bylo by proto jistě zajímavé provést výzkum adaptace na mluvčího v šepotu a určit, zda předchozí znalost projevu konkrétního mluvčího zlepší rozpoznatelnost jeho zamýšlené intonace.

Přínosem pro výzkum by mohla být i tvorba korpusu šeptané češtiny s různými typy šeptaných projevů. Mezi v současné době dostupnými šeptanými korpusy jsou například první anglickojazyčný korpus šeptané řeči wTIMIT používaný k trénování automatických rozpoznávačů, který vznikl v rámci disertační práce B. P. Lima (Lim, 2011), anglickojazyčný korpus CHAINS vytvořený za účelem výzkumu identity mluvčího, jehož součástí je i soubor různých typů šeptaných projevů (Cummins a kol., 2006), nebo srbský šeptaný korpus 50 izolovaných slov čtených 10 mluvčími v šeptané verzi a modální fonaci (Markovic a kol., 2013).

Dalším zajímavým podnětem může být i srovnání reálného šepotu a šepotu uměle vyrobeného z běžné fonace, které by mohlo napomoci odhalení případných kompenzačních mechanismů, které se v běžné řeči vyskytují v jiné míře než v řeči šeptané, a tudíž se jednoduchou manipulací signálu modální fonace (např. nahrazením hlasivkového buzení šumem) nedají převést do šepotu, který by působil přirozeně.

V neposlední řadě by bylo možné se zaměřit například na cizinecký přízvuk, který se projevuje jak v rovině hláskové, tak v rovině suprasegmentální. Suprasegmentální vlastnosti řeči značně přispívají k její přirozenosti jak při projevu v cizím jazyce, tak v syntéze řeči. Intonace v šepotu se nemusí nutně projevovat

ve všech jazycích stejně, šepot však zároveň často nebývá součástí výuky cizích jazyků. Nelze tedy bez experimentálního ověření snadno říci, zda bude člověk automaticky používat kompenzační mechanismy vlastního jazyka, či svůj projev v šepotu na základě projevu v daném jazyce v modální fonaci automaticky přizpůsobí jeho šeptané podobě.

Tato bakalářská práce přinesla základní vhled do podoby melodie v šeptané češtině, zdaleka však neobsáhla veškeré aspekty, které by bylo v této oblasti možné prozkoumat. Intonace je i v modální fonaci komplexním tématem a u intonace v typu fonace, který se vyznačuje nepřítomností základní frekvence, tomu rozhodně není jinak. Výzkum intonace v šeptané češtině tak skýtá nespočet dalších možností.

## 5. Závěr

V rámci této bakalářské práce jsem zkoumala rozpoznatelnost melodie v šepotu, ve kterém není přítomna základní frekvence jako hlavní akustické vodítko, a následně vliv zamýšlené melodie v šepotu na vybrané akustické parametry. K tomuto výzkumu jsem pořídila nahrávky slov se strukturou CVCV se znělými konsonanty realizované s různými melodickými průběhy, nejdříve formou opakování předem připravených nahrávek daných slov zasazených do vět a následně zpívané podle zvuku klavíru. Oba typy projevu byly nahrány v modální fonaci a v šepotu, předlohy byly vždy pro oba typy fonace stejné. Nahrávání se zúčastnily čtyři ženy se zkušenostmi se zpěvem.

Přibližně 80 % z každé skupiny šeptaných nahrávek jsem použila při tvorbě dvou online percepčních testů. První obsahoval slova zpívaná v šepotu, druhý nahrávky šeptaných slov z vět. Respondenti měli v obou případech určit, zda v melodii přehrané položky slyší stoupnutí, či klesnutí, a každou nahrávku si mohli jednou poslechnout znovu. Oba percepční testy vyplnilo stejných 33 respondentů. Výsledky ukázaly, že rozpoznání zamýšlené melodie v šepotu není vždy zcela dílem náhody, zároveň se také projevila variabilita v závislosti na mluvčí a poměru pomyslných základních frekvencí dvou tónů intervalu. Obecně byly v obou percepčních testech lépe rozpoznávány klesavé intervaly, zvláště u zpěvu pak vyššího podílu správných odpovědí dosahovaly položky, kde byl poměr zamýšlené základní frekvence dvou tónů v intervalu větší. Exploratorní pokus pracující s odpověďmi z úvodního dotazníku u percepčních testů naznačil, že by rozpoznání melodie v šepotu mohlo souviset se zkušenostmi s hudbou, pokus by však vyžadoval ověření experimentem zaměřeným na tuto problematiku.

V akustické analýze jsem pomocí lineárních smíšených modelů a následných testů poměru věrohodností analyzovala vliv zamýšlené melodie v šepotu na F1-F3, poměry těchto formantů, spektrální těžiště a spektrální sklon. Z formantů zamýšlená melodie u obou typů nahrávaného šeptaného materiálu statisticky významně

ovlivňovala druhý formant, u zpěvu v šepotu pak také F3 a poměr F2:F1 a F3:F2, který odráží větší pohyb F2 než prvního a třetího formantu. U zpěvu v šepotu se jako další parametry ovlivněné zamýšlenou melodií prokázaly spektrální těžiště a spektrální sklon.

U slov z šeptaných vět bylo kromě F2 zamýšlenou melodií ovlivněno už jen spektrální těžiště signálu filtrovaného tak, aby pásmo obsahující hlavní formantovou strukturu bylo odstraněno, což by mohlo naznačovat přítomnost akustických vodítek melodie v šepotu v pásmu vyšších frekvencí. Toto zjištění však vyžaduje další percepční experimenty a akustická měření jiné sady nahrávek, aby byla ověřena zobecnitelnost a praktická významnost výsledků a domněnka tak byla potvrzena či vyvrácena.

Tato práce přinesla zajímavá zjištění a srovnání vlastností šepotu v češtině s dosavadním výzkumem, zaměřeným především na holandštinu, francouzštinu a angličtinu. Závěry je však samozřejmě nutné ověřit dalšími experimenty a na jiných datech. Kromě cesty k ověření závěrů stávajícího experimentu práce také otevírá nové otázky v oblasti suprasegmentálních vlastností šepotu, nejen z hlediska melodie řeči.

# Seznam použité literatury

- APPLE INC (2004–2022). Garageband for Mac. Verze 10.4.6. <https://www.apple.com/mac/garageband/>.
- ASHBY, M. a MAIDMENT, J. (2016). *Úvod do obecné fonetiky*. Karolinum.
- AUDACITY TEAM (1999–2021). Audacity®. Verze 3.1.3. <https://audacityteam.org/>.
- BATES, D., MÄCHLER, M., BOLKER, B. a WALKER, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1–48. doi: 10.18637/jss.v067.i01.
- BOERSMA, P. a WEENINK, D. (1992–2021). Praat: doing phonetics by computer. Verze 6.1.56. <http://www.praat.org/>.
- BOERSMA, P. a WEENINK, D. (2001). Speak and unSpeak with Praat. *Glott international*, **5**, 341–347.
- BONNOT, J.-F. P. a CHEVRIE-MULLER, C. (1991). Some effects of shouted and whispered conditions on temporal organization. *Journal of Phonetics*, **19** (3), 473–483. doi: 10.1016/S0095-4470(19)30339-0.
- BOŘIL, T. (2017). Zvuková vlna. In KARLÍK, P., NEKULA, M. a PLESKALOVÁ, J., editors, *CzechEncy - Nový encyklopedický slovník češtiny*. Masarykova univerzita. URL <https://www.czechency.org/>.
- CIRILLO, J. (2004). Communication by unvoiced speech: the role of whispering. *Anais da Academia Brasileira de Ciências*, **76**(2), 413–423.
- COLEMAN, J., GRABE, E. a BRAUN, B. (2002). Larynx movements and intonation in whispered speech. Summary of research supported by British Academy grant SG-36269. <http://www.phon.ox.ac.uk/>.
- COTESCU, M., DRUGMAN, T., HUYBRECHTS, G., LORENZO-TRUEBA, J. a MOINET, A. (2020). Voice Conversion for Whispered Speech Synthesis. *IEEE Signal Processing Letters*, **27**, 186–190. doi: 10.1109/LSP.2019.2961213.
- CUMMINS, F., GRIMALDI, M., LEONARD, T. a SIMKO, J. (2006). The chains corpus: Characterizing individual speakers. In *Proceedings of SPECOM*, volume 6, pages 431–435.
- DONDERS, F. C. (1858). Ueber die Natur der Vocale. In *Archiv für die holländischen Beiträge zur Natur- und Heilkunde*. Bd.1. Kemink & Zoon.

- FÓNAGY, J. (1969). Accent et intonation dans la parole chuchotée. *Phonetica*, **20**(2-4), 177–192.
- GIET, F. (1950). *Zur Tonität nordchinesischer Mundarten*. Anthropos-Institut Sankt Augustin: Studia Instituti Anthropos. Verlag der Missionsdruckerei St. Gabriel.
- GIET, F. (1955). Kann man in einer tonsprache flüstern? *Lingua*, **5**, 372–381. doi: 10.1016/0024-3841(55)90029-2.
- HAAHR, M. (1998–2018). RANDOM.ORG: true random number service. <https://www.random.org>.
- HARRELL JR, F. E. (2022). *Hmisc: Harrell Miscellaneous*. Verze 4.7-0. URL <https://CRAN.R-project.org/package=Hmisc>.
- HEEREN, W. F. L. a VAN HEUVEN, V. J. (2014). The interaction of lexical and phrasal prosody in whispered speech. *The Journal of the Acoustical Society of America*, **136**(6), 3272–3289. doi: 10.1121/1.4901705.
- HEEREN, W. F. L. (2015). Vocalic correlates of pitch in whispered versus normal speech. *The Journal of the Acoustical Society of America*, **138**(6), 3800–3810. doi: 10.1121/1.4937762.
- HEEREN, W. F. L. a LORENZI, C. (2014). Perception of prosody in normal and whispered French. *The Journal of the Acoustical Society of America*, **135**(4), 2026–2040.
- HELMHOLTZ, H. (1863). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik: Mit in den Text eingedruckten Holzschnitten*. J. Vieweg.
- HIGASHIKAWA, M. a MINIFIE, F. D. (1999). Acoustical-perceptual correlates of “whisper pitch” in synthetically generated vowels. *Journal of Speech, Language, and Hearing Research*, **42**(3), 583–591. doi: 10.1044/jslhr.4203.583.
- ITO, T., TAKEDA, K. a ITAKURA, F. (2005). Analysis and recognition of whispered speech. *Speech Communication*, **45**(2), 139–152. doi: 10.1016/j.specom.2003.10.005.
- KLOSTER JENSEN, M. (1958). Recognition of word tones in whispered speech. *WORD*, **14**(2-3), 187–196. doi: 10.1080/00437956.1958.11659663.

- KUANG, J. a LIBERMAN, M. (2015). The effect of spectral slope on pitch perception. In *Sixteenth Annual Conference of the International Speech Communication Association*.
- LADEFOGED, P. a BROADBENT, D. E. (1957). Information Conveyed by Vowels. *The Journal of the Acoustical Society of America*, **29**(1), 98–104. doi: 10.1121/1.1908694.
- LENTH, R. V. (2022). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. URL <https://CRAN.R-project.org/package=emmeans>.
- LIM, B. P. (2011). *Computational Differences between Whispered and Non-whispered Speech*. PhD thesis, University of Illinois.
- LIU, S. a SAMUEL, A. G. (2004). Perception of mandarin lexical tones when f0 information is neutralized. *Language and Speech*, **47**(2), 109–138. doi: 10.1177/00238309040470020101.
- MACHAČ, P. a SKARNITZL, R. (2010). *Fonetická segmentace hlásek*. Epocha.
- MACHAČ, P. a ŠTURM, P. (2010). The phonological contrast of voicing in whispered Czech and its phonetic correlates – A preliminary study. *20th Czech-German Workshop - Speech Processing*, pages 34–43.
- MARKOVIC, B., JOVICIC, S., GALIC, J. a GROZDIĆ, D. (2013). Whispered speech database: design, processing and application. In *TSD 2013*. doi: 10.13140/2.1.4302.3366.
- MERTENS, P. (2004). Le prosogramme : une transcription semi-automatique de la prosodie. *Cahiers de l'Institut de Linguistique de Louvain*, **30**, 7–25. doi: 10.2143/CILL.30.1.519212.
- MEYER-EPPLER, W. (1957). Realization of Prosodic Features in Whispered Speech. *The Journal of the Acoustical Society of America*, **29**(1), 104–106. doi: 10.1121/1.1908631.
- NIEBUHR, O. (2009). Intonation segments and segmental intonation. In *10th Annual Conference of the International Speech Communication Association*, pages 2435–2438. ISCA. doi: 10.21437/Interspeech.2009-308.
- PALCOVÁ, Z., VEROŇKOVÁ, J., VOLÍN, J. a SKARNITZL, R. (2004). Stabilizace některých termínů pro fonetický popis češtiny v závislosti na nových výsledcích výzkumu. In *Sborník z Konference česko-slovenské pobočky ISPhS*, pages 65–74.

- PANCONCELLI-CALZIA, G. (1954). Das flüstern in seiner physio-pathologischen und linguistischen bedeutung. *Lingua*, 4, 369–378. doi: 10.1016/0024-3841(54)90071-6.
- PANCONCELLI-CALZIA, G. (1994). ‘Geschichtszahlen der Phonetik’ (1941), together with ‘Quellenatlas der Phonetik’ (1940): New edition, volume 16 of *Studies in the History of the Language Sciences*. John Benjamins Publishing Company. doi: 10.1075/sihols.16.
- PHAPATANABURI, K., PATHONSUWAN, W., WANG, L., ANCHUEN, P., JUMPHOO, T., BUAYAI, P., UTHANSAKUL, M. a UTHANSAKUL, P. (2022). Whispered speech detection using glottal flow-based features. *Symmetry*, 14 (4), 777. doi: 10.3390/sym14040777.
- POMPINO-MARSCHALL, B. (2003). *Einführung in die Phonetik*. De Gruyter. doi: 10.1515/9783110913248.
- PUFFER, H. (2018). *ABC des Sprechens: Grundlagen, Methoden, Übungen*. Henschel.
- R CORE TEAM (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- RAEESY, Z. (2018). Whisper to Alexa, and She’ll Whisper Back. <https://www.amazon.science/blog/whisper-to-alexa-and-shell-whisper-back>.
- RAEESY, Z., GILLESPIE, K., MA, C., DRUGMAN, T., GU, J., MAAS, R., RASTROW, A. a HOFFMEISTER, B. (2018). LSTM-based whisper detection. In *2018 IEEE Spoken Language Technology Workshop (SLT)*, pages 139–144. IEEE.
- ROMPORTL, M. (1973). *Základy fonetiky*. Státní pedagogické nakladatelství.
- SEGERBÄCK, B. (1965). La réalisation d’une opposition de tonèmes dans des dissyllabes chuchotés: étude de phonétique expérimentale. *Studia Linguistica*, 19(1-2), 1–54.
- SKARNITZL, R., ŠTURM, P. a MACHAČ, P. (2013). The phonological voicing contrast in Czech: an EPG study of phonated and whispered fricatives. In *14th Annual Conference of the International Speech Communication Association*, pages 3191–3195.



- SKARNITZL, R., ŠTURM, P. a VOLÍN, J. (2016). *Zvuková báze řečové komunikace: fonetický a fonologický popis řeči*. Univerzita Karlova v Praze, Nakladatelství Karolinum.
- STOET, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, **42**(4), 1096–1104. doi: 10.3758/BRM.42.4.1096.
- STOET, G. (2017). PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments. *Teaching of Psychology*, **44**(1), 24–31. doi: 10.1177/0098628316677643.
- SUITS, B. H. (1998–2022). Frequencies for equal-tempered scale, A4 = 440 Hz. <https://pages.mtu.edu/~suits/notefreqs.html>.
- SVATOŠOVÁ, M. (2021). Znělost vybraných párových hlásek v šeptané češtině. Bakalářská práce, Univerzita Karlova.
- THOMAS, I. B. (1969). Perceived Pitch of Whispered Vowels. *The Journal of the Acoustical Society of America*, **46**(2B), 468–470. doi: 10.1121/1.1911712.
- VOLÍN, J. (2010). Fonetika a fonologie. In CVRČEK, V., editor, *Mluvnice současné češtiny. 1: Jak se píše a jak se mluví*, pages 35–64. Nakladatelství Karolinum.
- WENNDT, S. J., CUPPLES, E. J. a FLOYD, R. M. (2002). A study on the classification of whispered and normally phonated speech. In *Seventh International Conference on Spoken Language Processing*.
- WICKHAM, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. URL <https://ggplot2.tidyverse.org>.
- ZHANG, C. a HANSEN, J. H. (2007). Analysis and classification of speech mode: whispered through shouted. In *Eighth Annual Conference of the International Speech Communication Association*.
- ŽYGIS, M., PAPE, D., KOENIG, L. L., JASKUŁA, M. a JESUS, L. M. (2017). Segmental cues to intonation of statements and polar questions in whispered, semi-whispered and normal speech modes. *Journal of Phonetics*, **63**, 53–74. doi: 10.1016/j.wocn.2017.04.001.