

---

UDK 808.50-48  
Izvorni znanstveni rad

---

Prihvaćeno 14.05.1998.

**Zdena Palková**  
Filozofski fakultet  
Sveučilište u Pragu  
Češka

**PERCEPČNÍ TOLERANCE PŘI POSUZOVÁNÍ NĚKTERÝCH  
INTONAČNÍCH JEVŮ V ČEŠTINĚ**

*PERCEPTIVNA SNOŠLJIVOST PRI PRIHVAĆANJU INTONACIJSKIH  
OBILJEŽJA U ČEŠKOM JEZIKU*

**SAŽETAK**

*Na svakoj razini fonetskih istraživanja u središtu je pozornosti odnos između zvukovnih ostvaraja pojedinih govornih tekstova i općih pojmova bitnih za opis zvukovne strukture jezika. Ključni korektivni činitelj ovdje je percepcija govora na koju se odnosi to veći broj pokusa u istraživanju govora što se više razvija instrumentalna i kompjuterska tehnologija. Primjerice, tvorba pravila za automatsku TTS sintezu često zahtijeva da se pojednostavi odnos među govornim pojavama, a uvijek uključuje smanjenje njihove neodređenosti. Ako pratimo*

---

snošljivost slušatelja prema odklonima od očekivanog oblika na materijalu čiji je sastav poznat i koji je u stanju podnijeti promatrane promjene, dobivamo značajne nam obavijesti i o važnosti individualnih zvukovnih parametara, i o udjelu neodređenosti u govornoj komunikaciji.

Iskustvo nam, međutim, govori da prosudbe dobivene preko perceptivnih testiranja treba tumačiti uz razlikovanje prirode zadatka koji se postavlja pred slušatelja. U istraživanju prozodije češkog jezika do sada smo razlikovali tri njezina oblika:

1. stupanj snošljivosti s obzirom na razumijevanje govora
2. stupanj snošljivosti s obzirom na ukupnu komunikativnu prihvatljivost govora
3. utjecaj nenamjernog davanja prednosti pri perceptivnoj ocjeni

Pod razumljivošću podrazumijevamo postizanje suglasnosti između govornika i slušatelja, tj. treba biti udovoljen uvjet da slušatelj razumije smisao upravo onako kako je to govornik htio.

Zamisao o komunikacijskoj prihvatljivosti nekad se povezivala najčešće sa govornim poremećajima i njihovim prevladavanjem. Danas se ona najčešće odnosi na sve veće nastojanje da se koristi stvoreni sintetizirani govorni signal. Osim razumljivosti, postoje zahtjevi i za olakšano percipiranje govora, a tu spadaju zahtjevi za raznolikošću oblika, za odgovarajućom brzinom i odgovarajućom reljefnosti intonacijskih obrazaca i drugo.

Kada se provode pokusi sa svrhom poboljšanja sinteze iznad neke osnovne razine razumljivosti može se dogoditi da se razumljivost govora i njegova prihvatljivost nužno ne podudaraju jedna s drugom. Primjerice u češkom jeziku se naglašava početni slog u riječi kako bi se pojačala njegova delimitativna funkcija, a posljedica tog obilježja jest slaba prihvatljivost za slušatelje, jer se taj način izgovora doživljava kao previše stereotipni. Također, oštro razlikovanje uzlazne intonacije upitne rečenice koja počinje upitnom riječju od uzlazne intonacije rečenica koje očekuju odgovor sa da ili ne, zahtijeva značajne visinske skokove, a negativno je ocijenjen od slušatelja. Čini se da slušatelj više voli slušni dojam koji je što sličniji onom što ga ostavlja prirodni govor, čak i ako to znači stanovitu neodređenost koja bez svake sumnje postoji i u normalnoj komunikaciji. Treći oblik koji valja uzeti u obzir, odnosi se na ono što slušatelji vole. Ponekad u slušnim tekstovima nađemo da slušatelji preferiraju zvukovnu varijantu među opcijama bez prividne motivacije, primjerice iz ne-zvukovnog konteksta ili iz uvjeta pokusa. Takve slušne preferencije unose izobličenje u rezultate procjene, te (s obzirom na uporabu slušnih testova kojima se procjenjuje važnost govornih pojava) možda predstavljaju i zaseban metodološki problem.

Članak daje nekoliko primjera podataka iz niza istraživanja na češkom koja se bave obilježjima zvukovne jedinice na razini riječi. Istraživanje uključuje slušne testove koji se sastoje od nizova slogova čije značenje ovisi o načinu segmentiranja u riječi (ili fonetskoj riječi, da se poslužimo strožom terminologijom), a o čemu odlučuje ispitanik. Znači, ispitanici donose odluke na osnovi svojeg jezičnog iskustva, a ne na osnovi teorijskih pretpostavki.

(primjerice: "ale-VČE-RA-TO-PI-LI dobrovolne", što znači "ali jučer su XXX dobrovoljno", gdje se XXX može zamijeniti ili sa TO PILI ili sa TOPILI.) Ovi su testovi rađeni i s prirodnim i sa sintetskim govorom.

Za slijed s više od pet slogova u materijalu prirodnoga govora, stupanj slaganja između govornikove namjere i slušateljeve procjene iznosi otprilike 70-75 %. Test na sintetskom govoru mijenja rezultat utječući na Fo (nulti formant). Testovi koji se koriste nizom intonacijskih modela iz materijala koji su slušatelji ocijenili primjenjivima na čitanje neutralnih tekstova, posjeduju očividno niži stupanj slaganja između očekivanja i slušateljeve procjene nego kada se radi o prirodnim podacima (otprilike za 55-60 %).

Međutim, svi materijali pokazuju istu tendenciju, te preferiraju segmentaciju petosložnog niza u lanac 3:2 (vidi tabelu A), a ne 2:3 (vidi tabelu B). Pokus s prirodnim govorom nosi naziv test Pri 1 na tabelama, a pokusi sa sintetskim govorom test Bele.

Tabela 1 pokazuje u postocima ocjene koje se slažu s izvornom namjerom govornika. Taj postupak ne uzima u obzir da li je rečenicu "priznao" veći ili manji broj slušatelja.

Tabela 2 (za signale prirodnoga govora) kao i tabela 3 (za TTS) daje detaljnije rezultate, a postupak je osjetljiviji na procjenu većine u pojedinačnim testovima. Očito je da postoji veće slaganje slušatelja za varijantu 3:2, bez obzira na to radi li se o namjeri govornika ili o TTS (vidi stupac A, red A+a nasuprot stupac B, red B+b, na tabelama 2 i 3). Osim toga, test te vrste pokazuje preference ovisne o tekstu testa, posebno glede leksika upotrebljenih konstrukcija. Na sintezu mogu istodobno djelovati sitne razlike koje ovise o kombinaciji glasova, naročito nepravilnosti koje nastaju tijekom prijelaza. Tabela 4 pokazuje tu vrstu preference na testu Bele u kojem tri rečenice imaju različito tumačenje, tri teže segmentaciji tipa 3:2, a tri prema tipu 2:3 (vidi tabelu 4). Opet je karakteristično da je ova leksički ovisna većina slabija kod segmentacije tipa 2:3 nego 3:2.

Tabela 4 nadalje pokazuje jaku tendenciju lomljenja testiranog slijeda slogova prema Fo tipu obrisa (4), što se procjenjuje kod tipa 3:2 koji prevladava u svim dijelovima pokusa. Činjenica da on vjerojatno prevladava nad različitim vrstama preferenci potvrđuje istaknutost te tendencije.

Jedan od glavnih ciljeva ovog istraživanja češke prozodije jest traženje zvukovnih karakteristika za koje pouzdano možemo reći da prednjače u percepciji govora. S te točke gledišta rezultati dobiveni primjenom raznih obilježja - primjerice intonacijskih obrisa - u kontekstu poznatih preferencija mogli bi dati kriterij za tu pouzdanost.

Što bolje uvjete tehnologija pruža za istraživanje zacrtanog niza problema, to više u prvi plan dolazi opće pitanje koje se odnosi na udio pojedinačnih činitelja koji su zajedno odgovorni za razumljivost u pravom govornom okruženju. U tom pogledu treba posebno proučavati uzajamne odnose među zvukovnim činiteljima, nezvukovnim komponentama jezične konstrukcije, posebno one semantičke, te izvanlingvističke situacijske činitelje.

*Teško je sumnjati da različite razine snošljivosti onih koji slušaju svoj izvorni jezik i procjenjuju pojedina zvukovna obilježja govora imaju veze s jezičnim sustavom. To omogućuje usporedna istraživanja, posebice u srodnim jezicima.*

***Ključne riječi: intonacije, komunikacijska prihvatljivost, češki jezik***



## 1. 0

Vztah mezi zvukovými jevy konkrétních řečových textů a zobecňujícími pojmy, jejichž prostřednictvím je vytvářen strukturní popis zvukové stavby jazyka, je trvalým předmětem pozornosti v každé fázi fonetického výzkumu. Teoretické pojmy jsou vymezovány jako statické, diskretní, vágnosti prosté skladební prvky, zvuk řeči je kontinuální multidimenzionální materie s proměnlivou valencí navzájem se ovlivňujících složek. Odlišit zobecňování od hypostaze někdy není snadné.

Účinným korektivním faktorem může být percepce řeči, jejíž projevy lze s rozvojem instrumentální a počítačové techniky zachycovat stále spolehlivěji. Tak například formulace pravidel pro automatickou syntézu text-to-speech vyžaduje občas zjednodušení vztahů mezi jazykovými jevy a vždy znamená oslabení či odstranění jejich neurčitosti. Sledování tolerance posluchače vůči odchylkám od předpokládaných forem, na materiálu, jehož složení známe a můžeme řízeným způsobem obměňovat, přináší pozoruhodné informace, jak o váze jednotlivých zvukových parametrů, tak o podílu vágnosti v řečové komunikaci.

## 2.0

Přitom se ukazuje, že hodnotící soudy získané poslechovými testy je nutno interpretovat diferencovaně, podle povahy úkolu, který posluchač řešil. Při výzkumu prozodie češtiny zatím odlišujeme trojí aspekt:

- stupeň tolerance ve vztahu k srozumitelnosti projevu,
- stupeň tolerance z hlediska celkové komunikační přijatelnosti projevu,
- přítomnost nezáměrných preferencí v percepčním hodnocení.

2.1 Pojmem *srozumitelnost* se přitom míní dosažení shody v interpretaci řečeného mezi mluvčím a posluchačem, tedy stav, kdy posluchač „rozumí“ textu právě v tom významu, jaký zamýšlel mluvčí.

Problematika srozumitelnosti v tomto pojetí zahrnuje i určování stupně jednoznačnosti interpretace. Na tom bývají založeny experimenty, jejichž pomocí zjišťujeme platnost určitých zvukových prostředků v určité funkci. Klasickým příkladem jsou testy, v nichž posluchač vybírá jeden z alternativních významů, přičemž zvukový prostředek je jediným diferencujícím prvkem mezi nabízenými variantami textu.

2.2 *Komunikační přijatelnost* je pojem, který se v minulosti vázal spíše k problematice poruch řeči a jejich překonávání. Jeho aktualizace v současné době souvisí především s rostoucí snahou využívat v různých odvětvích praxe automaticky vytvářený syntetický řečový signál. Nezbytným předpokladem bývá požadavek, aby výslednou podobu řeči byl laický posluchač ochoten akceptovat jako dorozumivací prostředek i v delším časovém úseku. Nároky jsou v tom případě kladeny nejen na srozumitelnost, ale také na snadnou vnímatelnost řeči,

což zahrnuje takové požadavky, jako je adekvátnost tempa, úměrná výraznost intonačních forem, dostatečná variabilita zvukové podoby, atd.

Při pokusech o zlepšování syntézy, která dosáhla základní úrovně srozumitelnosti, zjišťujeme, že tato dvě hlediska, srozumitelnost a přijatelnost řeči, nepůsobí vždy ve vzájemném souladu. Např. v češtině při zařazení zvukové charakteristiky, která zdůrazňuje první slabiku slova a tím posiluje (často citovanou) delimitativní funkci, získáváme projev, který posluchači považují za málo přijatelný, protože je stereotypní a unavuje je. (srov. koncepce Palková - Ptáček 1994 a nová varianta Palková - Ptáček 1995, 1997, či Palková 1997). Podobně zřetelné odlišení stoupavé melodické kadence větné pro zjišťovací otázku a stoupavé kadence pro neukončenou výpověď vyžaduje použít výrazné výškové kroky a opět mívá za následek negativní hodnocení posluchačů. Ždá se, že posluchač při poslechu syntézy preferuje zvukový dojem, který se co nejvíce blíží přirozené řeči, a to i za cenu jistého podílu neurčitosti, která ovšem existuje i v běžné komunikaci.

2.3 Třetí aspekt, který je třeba brát v úvahu, nazýváme zde *poslechové preference*. V poslechových testech se občas setkáváme s tím, že posluchači při výběru z nabízených možností preferují určitou zvukovou variantu, aniž by existovala motivace např. z nezvukového kontextu či z podmínek samotného experimentu. Tyto poslechové preference vnašejí do výsledků hodnocení určitý druh zkreslení a z hlediska využívání poslechových testů pro posuzování relevance zvukových jevů představují samostatný metodický problém.

### 3.0

Jako příklad uvedeme jeden okruh výzkumu češtiny, v němž se v různé míře projevují výše naznačené metodické problémy. Výzkum sám probíhá řadu let a je zaměřen k zjišťování vlastností zvukové jednotky na úrovni slova. Dosavadní výsledky umožňují formulovat dobře přijatelnou podobu prozodie v automatické syntéze, ale nejsou např. dostačující k formulaci konkrétních hypotéz pro usnadnění automatického rozpoznávání souvislé řeči.<sup>1</sup>

3.1 Částí tohoto výzkumu jsou poslechové testy, v nichž jsou pokusným osobám předkládány k rozhodnutí sekvence slabik, jejichž význam se může měnit podle jejich různého rozčlenění do slov, či přesněji do přízvukových taktů. Posluchači tedy rozhodují na základě své jazykové zkušenosti, nikoli na základě teoretických předpokladů o podobě zvukové jednotky. Tyto testy byly postupně použity na materiálu z řeči přirozené i syntetické.

Materiál z přirozené řeči získáváme tak, že uvedené páry jsou zařazeny do souvislého textu, přičemž stojí vždy v mediální pozici věty. Texty jsou dostatečně dlouhé, takže si mluvčí při jejich čtení neuvědomují přítomnost dvojznačných sekvencí. Z takto získaných projevů jsou pak tyto sekvence izolovány a předkládány posluchačům k rozhodnutí, mohou být analyzovány atd.

<sup>1</sup> Výzkum, z něhož je citováno, má podporu grantu GAČR 405/96/0301

Syntetický materiál obsahuje testované dvojice obvykle v kontextu ukončené věty jejíž rytmická struktura zůstává beze změny. Předepsaným způsobem je modifikována některá zvuková kvalita, zatím ponejvíce průběh F0. Celé věty jsou pak předkládány posluchačům k rozhodnutí.

Dále uvedené příklady jsou částí souboru, v němž je testována variabilita členění pětislabičné sekvence na posloupnost taktů se slabičnou délkou 3:2 a 2:3. Příkladem mohou být dvojice:

TO PILI vs. TOPILI („Ale VČE - RA - TO - PI - LI dobrovolně.“)  
 SE LHALO vs. SELHALO („Ale A - SI - SE - LHA - LO několikrát“)  
 JÍ DALI vs. JÍDALI („Ale TEN - KRÁT - JÍ - DA - LI pomeranče“)

Celkově lze říci, že výše shody mezi záměrem mluvčího a hodnocením posluchače pro uvedenou pětislabičnou sekvenci se v přirozeném řečovém materiálu pohybuje přibližně v pásmu 70-75%. V testu na bázi syntézy závisí výsledek na tom, jak výrazné změny F0 jsou používány. V testech, v nichž je aplikován výběr intonačních kontur ze souboru, který byl posluchači označen za vhodný pro čtení souvislých neutrálních textů, je shoda mezi předpokladem a hodnocením posluchače zřetelně nižší, než v přirozeném materiálu, přibližně 55-60 %.

Ve všech materiálech se však objevuje stejná tendence k preferenci členění na slabičnou posloupnost 3:2 (dále A) před posloupností 2:3 (dále B).

3.2 Jako příklad uvedeme výsledky dvou testů obsahující shodné sekvence slov. V testu PRI-1 je materiálem přirozený řečový signál, v testu BELE signál porizovaný syntézou.

a)

Tab. 1. uvádí v % objem všech soudů, které se shodovaly s původní intencí mluvčího. Při tomto zpracování se nepřihlíží k tomu, zda daná věta byla „poznána“ větším nebo menším počtem posluchačů.

**Tabulka 1. Shoda soudů posluchačů s předpokladem (celkový počet soudů):**

	Realizace A + soudy A	Realizace B + soudy B	Celkově
Test PRI 1	80%	65%	73%
Test BELE	65%	52%	58%

Komentář k materiálu v obou testech je uveden dále u tab. 2 a 3.

b)

Tab. 2. (pro přirozený řečový signál) a tab. 3. (pro materiál syntézy) ukazují výsledky těchto testů podrobněji, přičemž zpracování bere v úvahu převahu



hodnocení v jednotlivých položkách testu. V obou souborech se opět projevuje preference varianty 3:2. Svědčí o ní zřetelně vyšší shoda posluchačských soudů se záměrem mluvčího, případně se zadáním syntézy (viz v tab. 2 a 3 vždy sl. A, ř. A+a proti sl. B, ř. B+b).

**Tabulka 2. Shoda soudů posluchačů s předpokladem - Test PŘI 1**  
(na základě rozhodování o jednotlivých položkách testu)

Hodnocení (posluchač)	Typ realizace (mluvčí)					
	A=3:2	B=2:3	Σ	A%	B%	Σ%
A	22	-				
a	4	5				
ΣA+a	26	5	31	84	16	50
?	3	8	11	10	26	18
B	1	14				
b	1	4				
ΣB+b	2	18	20	6	58	32
ΣΣ	31	31	62	100	100	100

Komentář k tab.2.

a) Interpretace shody v hodnocení na základě soudů „a“ v %:

100-75 = A, 74-60 = a, 59-41 = ?, 40-26 = b, 25-0 = B

b) Materiál: přirozený řečový signál. Rozsah: 11 sekvencí pěti slabik, v nichž o významu rozhoduje taktová hranice 3:2 nebo 2:3 (př. *včera to / pili* vs. *včera / topili*). Testované položky byly čteny ve větách zařazených v souvislém textu.

3 mluvčí, 4 položky vyloučeny pro technickou chybu, použito celkem 31 dvojic.

60 posluchačů.



**Tabulka 3. Shoda soudů posluchačů s předpokladem - Test BELE**  
(na základě rozhodování o jednotlivých položkách testu)

Hodnocení (posluchač)	A=3:2 B=2:3		Σ	Typ realizace (syntéza)		Σ%
	A%	B%		A%	B%	
A	19	9				
a	3	5				
ΣA+a	22	14	36	61	39	50
?	5	6	11	14	17	15
B	2	7				
b	7	9				
ΣB+b	9	16	25	25	44	35
ΣΣ	31	31	62	100	100	100

Komentář k tab. 3.

a) Interpretace shody v hodnocení na základě soudů „a“ v %:

$$100-75 = A, 74-60 = a, 59-41 = ?, 40-26 = b, 25-0 = B$$

b) Materiál: difonová syntéza. Rozsah: 9 sekvencí pěti slabik, v nichž o významu rozhoduje taktová hranice 3:2 nebo 2:3 (př. *včera to / pili* vs. *včera / topili*). Testované položky byly čteny ve větách se strukturou  $[2 + (2 / 3 + 3 / 2) + 4]$ .

Použity byly 4 různé varianty F0 u každé dvojice, celkem 36 dvojic. 33 posluchačů.

3.3 Kromě zmíněné tendence k preferenci jednoho z typů se ve všech testech tohoto typu objevují případy, v nichž je výsledné hodnocení posluchačů shodně v protikladu se záměrem mluvčího (v tab. 2. a 3. vždy sl. A, ř. B+b. a sl. B, ř. A+a). Soudíme, že je to dokladem o významu delšího zvukového kontextu při rozhodování posluchače (Palková 1996). I v těchto případech se však ukazuje převaha variant s rytmickou strukturou 3:2.

3.4 Při testech tohoto typu je dále třeba počítat s preferencemi, jež jsou závislé na textu testovaných položek, zejména z hlediska lexikálního obsazení použitých konstrukcí. V syntéze se současně s tím může projevit i vliv drobných diferencí závislých na kombinatorice hlásek, zejména nepravidelností v rovině trvání tranzientních fází. V našem materiálu se objevují určité preference tohoto typu v testu BELE, v němž 3 věty byly interpretovány proměnlivě, 3 se sklonem k členění 3:2, a 3 se sklonem k členění 2:3, viz tab. 4. Je příznačné, že i tato, na lexikálním obsazení závislá převaha je slabší u typu členění 2:3 proti typu 3:2.

**Tabulka 4. Test BELE - Shoda hodnocení s předpokladem (%) při rozložení podle intonačních typů a skupin vět**

		% soudů A+a u typu A (3:2)					% soudů B+b u typu B (2:3)				
varianta F0		1	2	3	4	S1-4	5	6	7	8	S5-8
		\\	V	Λ	//		\\	V	Λ	//	
věta	3+5+8	50	51	67	88	64	50	55	50	53	52
(var.)											
věta	1+7+9	85	84	81	89	85	40	25	20	14	25
(+3:2)											
věta	2+4+6	32	28	33	84	44	81	78	78	76	79
(+2:3)											
Σ	1-9	66	54	60	87	65	57	53	49	48	52

Komentář k tab. 4:

a) varianty F0 obsahují čtyři kombinace poklesu a stoupenutí F0 v testovaných dvojicích taktů, viz třetí řádek tabulky

b) testované dvojice slovních spojení

ve větách 3+5+8: v ní mají / vnímají; se lhalo / selhalo; jí dali / jídali

ve větách 1+7+9: to pili / topili; se strojí / sestrojí; se stavil / sestavil

ve větách 2+4+6: se dává / sedává; s ní žili / snížili; by dleli / bydleli

3.5 Tab. 4. zároveň ukazuje silnou tendenci k rozčlenění testované sekvence slabik v závislosti na typu kontury F0 (4), která je ve všech částech experimentu hodnocena s převahou jednoho typu (3:2). Skutečnost, že zřejmě převažuje nad preferencemi různého druhu, svědčí o výraznosti této tendence.

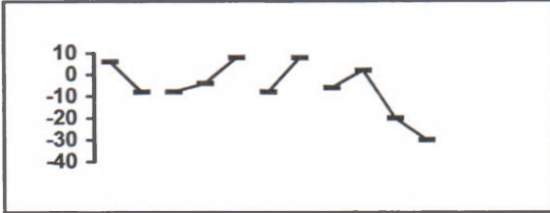
Zjišťování zvukových vlastností, které se prosazují při percepci řeči s dostatečnou spolehlivostí, je jedním z hlavních cílů současného výzkumu české prozodie. Má význam mimo jiné pro výběr zvukových charakteristik použitelných v projektu automatického rozpoznávání souvislé řeči. Výsledek aplikace různých charakteristik, zde např. intonačních kontur, v prostředí známých preferencí může z tohoto hlediska poskytnout určité kritérium této výraznosti.

3.6 Výzkum pomocí syntézy umožňuje sledovat také účinek změn ve velikosti výškových kroků u jednotlivých kontur F0. Jako příklad uvedeme tři ukázky intonačních kontur použitých v materiálu syntézy. Ve všech variantách je intonace modelována jako diference F0 jednotlivých slabik od relační hladiny 100% (blíže Palková - Ptaček 1997, Palková 1997).

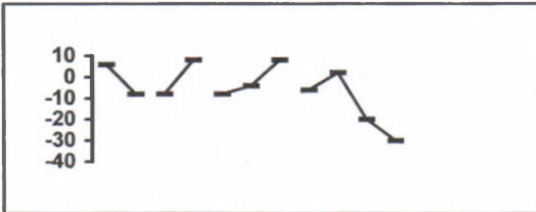
Příklad A1 - A2 ukazuje výrazné intonační kontury, které vedou v každém textu k rozhodnutí o členění pětislabičného řetězu ve shodě s původním záměrem na úrovni shody posluchačů větší než 75%.

**Příklad A1**

+6-8/-8-4+8/-8+8/-6+2-20-30

**Příklad A2**

+6-8/-8+8/-8-4+8/-6+2-20-30

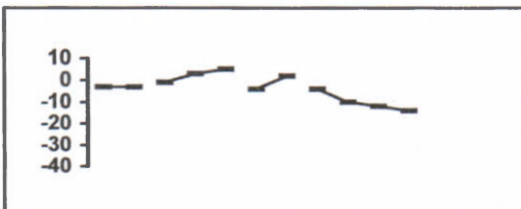


Příklady B a C ukazují dvě kontury použité v testu BELE. V tomto testu byly ověřovány kontury s minimální velikostí kroků F0. V příkladech B a C vždy varianta B1 a C1 člení sekvenci podle předpokladu, varianta B2 a C2 ukazuje podobu protikladné interpretaci.

Příklad B1- B2 ukazuje konturu F0-4 (viz tab.4.).

**Příklad B1:**

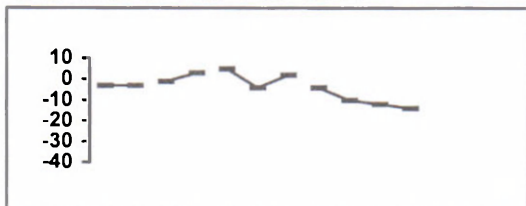
-3-3/-1+3+5/-4+2/-4-10-12-14





Príklad B2:

-3-3/-1+3+5/-4+2/-4-10-12-14



Príklad C1-C2 ukazuje konturu F0-1, v níž jsou použity nejméně výrazné intonační změny a rozhodnutí je patrně nejvíce ovlivněno sekundárními preferencemi.

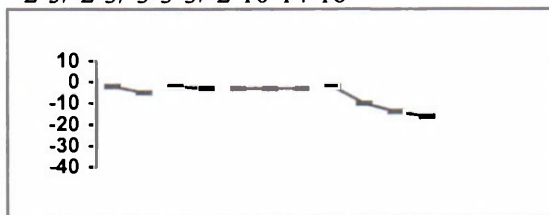
Príklad C1:

-2-5/-2-3-3/-3-3/-2-10-14-16



Príklad C2:

-2-5/-2-3/-3-3/-2-10-14-16



#### 4.0 Závěrem

Znalost různých preferencí posluchačů umožňuje reálněji interpretovat získané výsledky a také vytvářet testy, které se vyhnou některým zkreslením.

Cím lepší jsou technické možnosti pro zkoumání výše naznačených problémových okruhů, tím výrazněji vystupuje do popředí obecná otázka, jaký je vlastně podíl jednotlivých faktorů zúčastněných na dosažení srozumitelnosti v reálné řečové situaci. V daných souvislostech jde zejména o vztahy mezi

---

zvukovými faktory, nezvukovými, zejména sémantickými složkami jazykové stavby a mimojazykovými faktory situačními.

Sotva lze pochybovat o tom, že různý stupeň tolerance rodilých posluchačů při posuzování jednotlivých zvukových kvalit řeči má vazbu k systému jazyka. Pro srovnávací výzkum (zejména příbuzných jazyků) se tak otvírá prostor, který je zatím jen málo zaplněn.

#### LITERATURA

- Palková, Z., M. Ptáček** (1994) : Ein Beitrag zur Intonation in der Diphonsynthese . In: *Phonetica Pragensia VIII., AUC Philologica 1*, P. Janota, O. Schulzová (eds.), Praha, pp. 59-72
- Palková, Z., M. Ptáček** (1995): Prosodic issues in Czech: an application in TTS. In: *Proceedings of the XIIIth ICPhS, Stockholm, Vol.4*, pp. 380-383.
- Palková, Z.** (1996): Concurrent context intentions: a relevant factor in sound structures. In: *Phonetica Pragensia IX, Charisteria Premysl Janota oblata, AUC, Philologica 1*, Z. Palková (ed.), Praha, pp. 207-216
- Palková, Z.** (1997): Suprasegmentální vlastnosti češtiny v současném pohledu. In : *Sborník referátů ze zasedání Mezinárodní komise pro fonetiku a fonologii slovanských jazyků, Prešov 1996*, v tisku
- Palková, Z., M. Ptáček** (1997): Modelling prosody in TTS diphone synthesis in Czech. In: *Speech processing* , H.-W. Wodarz (ed.), *Forum Phonicum 63*, Frankfurt a.M., pp. 59-77.
-

**Zdena Palková**  
Faculty of philosophy  
University of Prague,  
Czech Republic

## PERCEPTIVE TOLERANCE IN ACCEPTABILITY OF INTONATION FEATURES IN CZECH

### SUMMARY

*The relationship between sound phenomena of particular speech texts and generalised concepts instrumental in description of sound structure of language is in focus at any level of phonetic research. The relevant corrective factor is speech perception whose research obtains growing experimental support as the instrumental and computational technology develops. For instance, rule formation for automatic TTS synthesis often demands simplification of the relationships among speech phenomena and always entails reduction of their vagueness. Monitoring tolerance of listeners for deflections from expected forms on material whose composition is known and capable of yielding to controlled change, brings out remarkable information both on weight of individual sound parameters and the share of vagueness in speech communication.*

*The experience, however, shows that evaluative judgements obtained through perceptive testing should be interpreted with distinctions as to the nature of the task dealt with by the listener. In researching the prosody of Czech, three aspects are distinguished so far:*

- *a degree of tolerance related to the intelligibility of the speech*
- *a degree of tolerance related to the overall communicative acceptability of the speech*
- *an influence of unintended preferences in the perceptive evaluation.*

*By intelligibility is meant reaching agreement in the interpretation of the text between the speaker and the listener, i. e. the condition when the listener 'understands' the text in just that meaning as intended by the speaker.*

*The idea of communicative acceptability used to link most often to issues of speech disturbances and their overcoming. At present, it relates mostly to the*

---



increasing effort to use machine-generated synthetic speech signal. Apart from intelligibility, demands are also made on easy perceptibility of speech, which includes requirements as sufficient variability of forms, adequacy in rate, adequacy in salience of intonation patterns, etc.

While experimenting to improve the synthesis beyond a certain basic level of intelligibility, we find the aspects of speech intelligibility and acceptability may then not necessarily always agree with each other. For instance the sound characteristics in Czech of pointing up the word-initial syllable to reinforce the delimitative function produces a manner of speech with low acceptability for the listeners due to being felt too stereotypical. Similarly, a sharp distinction of a rising melody sentence pattern in *wh*-questions from a rising pattern in open-end utterances calls for using conspicuous jumps in pitch, but again results in negative judgements with listeners. The listener seems to prefer a sound impression in listening to synthesis that gets as close as possible to natural speech, even if certain vagueness is traded off, which no doubt is there in ordinary communication as well.

The third aspect to take into consideration we call listening preferences. Now and then we encounter in listening tests that listeners prefer a sound variant out of the options with no apparent motivation e.g. from non-sound context or the conditions of the experiment. Such listening preferences bring a distortion to the evaluation results, and, considering the use of listening tests to evaluate the relevance of speech phenomena, possibly make a separate methodological issue.

The contribution provides a few examples of data from a range of research in Czech devoted to finding out features of a word-level sound unit. The research includes listening tests which provide sequences of syllables for the tested subjects to decide while the meaning depends on the way the sequences are segmented to make words (or stress units in stricter terminology). Hence the listeners decide drawing on their language experience, not on theoretic assumptions. (E.g. „ale VCE-RA-TO-PI-LI dobrovolně“ where either TO PILI or TOPILI is to be identified - literally „but YESTERDAY XXX voluntarily“ with either THEY DRANK IT. or THEY HEATED to replace the XXX.) These test were made out of both natural and synthetic speech.

The degree of agreement between the speaker's original intention and the listener's evaluation seems to be within roughly 70-75% range for the above 5-syllable sequence in natural speech material. A synthesised speech test influences the result with how conspicuous changes in  $F_0$  are applied. The tests which apply a set of intonation patterns out of a collection that was identified on continuous texts by listeners as applicable for reading neutral texts have an obviously lower degree of agreement between the expectation and the listener's evaluation than in natural data, approximately 55-60%.

All materials, however, show the same tendency to prefer the 5-syllable sequence segmentation into a 3:2 chain of syllables (see tab. A) over a 2:3 chain

(see tab. B). The natural speech experiment bears the label *Test Při 1* in the tables while the synthetic speech experiment is *Test BELE*.

The Table. 1 shows the amount in % of all the judgements in agreement with the speaker's original intention. This procedure does not take into consideration whether a sentence was 'recognised' by a greater or lesser number of the listeners.

The Table. 2 (for natural speech signal) as well as the Table. 3 (for TTS) offer results in a greater detail while the procedure is sensitive for majority evaluation in single test items. The 3:2 variant preference is made manifest by greater agreement in listeners' judgement with the speaker's intention or the TTS command, respectively (see col. A row A+a against col. B row B+b in tables 2 and 3).

Moreover, the test of this type show preferences depending on the text of tested items, notably concerning the lexical cast of the constructions used. Synthesis can simultaneously get affected with an influence of minute differences depending on sound combinations, especially irregularities during the course of transitions. The Table. 4 displays this type of preferences in the BELE test, where 3 sentences have varied interpretation, 3 tend to 3:2-type segmentation while 3 inclines towards the 2:3 type, see the Table. 4. Characteristically again, this lexically dependent majority is weaker with the 2:3-type segmentation than the 3:2 type.

The tab. 4 shows additionally a strong tendency to break out the tested syllabic sequence according to an F0 type of contour (4), which is evaluated with one type (3:2) prevailing throughout all the portions of the experiments. The fact that it probably prevails over preferences of various kinds attests to the prominence of this tendency.

Searching for sound features that gain ground in speech perception with sufficient reliability is one of the chief aims in the current research of Czech prosody. From this viewpoint, the result obtained out of applying various characteristics, e.g. intonation contours as herein, within an environment of known preferences, could offer a criterion of that reliability.

The better opportunities technology provides for research of the outlined range of issues, the more into the foreground rises a general question what in fact is the share of individual factors working together to achieve intelligibility in a real speech setting. In this respect, the mutual relationships among sound factors, non-sound components of the language structure, particularly semantic ones, and extra-linguistic situational factors have to be studied in particular.

It can be hardly doubted that the various degree of tolerance with native listeners in evaluating individual sound qualities of speech is linked to the system of language. It seems to offer possibilities for comparative research specifically in related languages.

**Key words:** intonation, communicative acceptability, Czech language