A PERCEPTUAL AND ACOUSTIC STUDY OF MELODY IN WHISPERED CZECH WORDS

Adléta Hanžlová, Tomáš Bořil

Institute of Phonetics, Charles University in Prague

adleta.hanzlova@gmail.com, tomas.boril@ff.cuni.cz

Introduction

The perception of melody in speech depends mainly on the fundamental frequency (f_0) which reflects vocal fold oscillation speed. Whisper is defined by the absence of phonation and therefore the lack of f_0. Intended melody in whisper, however, seems to be discernible regardless.

Aim of the study

- perception experiment assessing the discernibility of melody in whispered Czech words and words sung in whisper
- acoustical analysis of the effect of intended melody in whisper on possible correlates

Possible correlates of melody in whisper

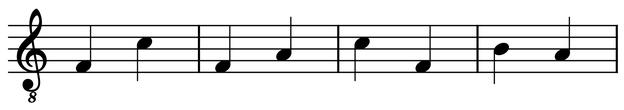
- formant frequencies [1–4]
- formant to formant ratios [4, 5]
- center of gravity (CoG) [6]
- spectral slope [5]

Material

4 female native Czech speakers aged 20-24 with musical education and experience with solo/choral singing

whispered speech: shadowing task

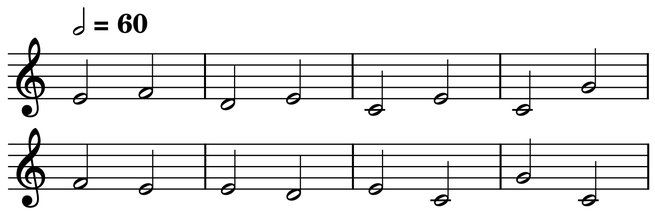
- 2-syllable onomatopoeic words [ba:ba, jɛ:jɛ:, la:la:, jo:jo:] set in the sentence "Řekl [target] anebo [modified target]," modified target stands for the target word with two short vowels.
- template sentences in modal phonation, melodic contours of target words manipulated in Praat [7] for f_0 of vowels to match musical intervals
- each sentence realized with 4 different melodic contours as statement/question (did he say [target] or [mod. target] / he said [target] or [mod. target])



Musical intervals melodic contours of template target words were manipulated to match.

words sung in whisper

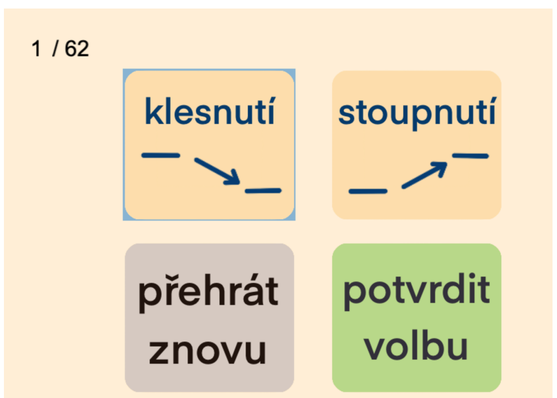
- 2-syllable onomatopoeic words [la:la:, jo:jo:]
- each target word realized with 8 different melodic contours matched to musical intervals
- piano track playing target intervals with metronome through headphones, each interval twice – speakers first listened, then sung along



Musical intervals played as template when recording sung and whisper-sung words.

Perception experiment

experiment setup:
→ 2AFC (fall/rise in melody)
→ 2 consecutive experiments
stimuli:
→ whispered target words
→ preceded by beep, 1 replay
→ words sung in whisper: 53 + 9 filler
→ whispered speech: 54 + 12 filler
respondents:
→ 33 Czech/Slovak aged 17–63, 13 male
analysis:
→ success rates for subgroups, confidence intervals with Bonferroni correction



Perception experiment interface
(buttons: fall, rise, replay, confirm choice)

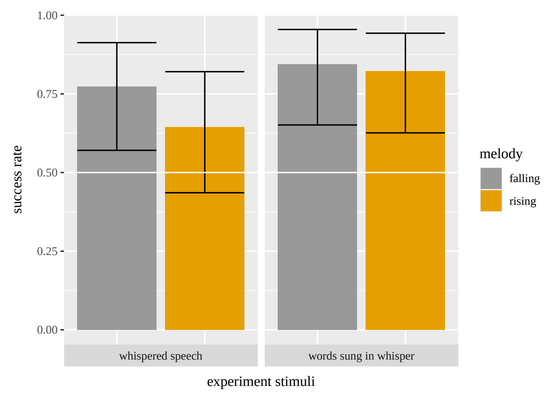
Acoustical analysis

middle third of vowels in target words
analysis in Praat [7], linear mixed-effects models in R [8], likelihood ratio tests
dependent variables (one per model):
→ formants (F1–F3)
→ formant ratios (F2:F1, F3:F1, F3:F2)
→ CoG, spectral slope

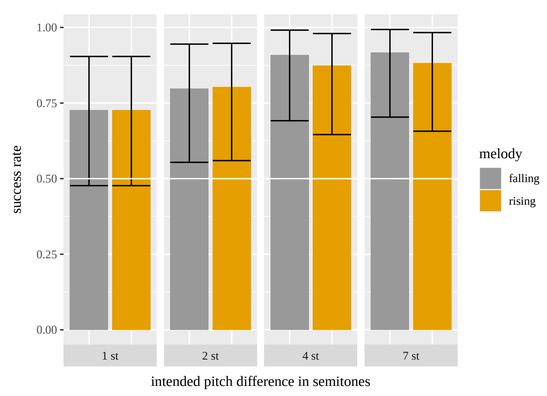
fixed effects (for all models):
pitch movement of target word (rise/fall) in interaction with position of vowel in the word (first/second syllable)

random intercepts (for all models):
→ speaker, target word

Discernibility of melody in whisper



Success rates of discerning melody in perception experiments by intended melody, confidence intervals at $\alpha = 0.05$, Bonferroni correction for $n = 2$.



Success rates of discerning melody of whisper-sung words by intended melody & pitch difference, confidence intervals at $\alpha = 0.05$, Bonferroni correction for $n = 8$.

Acoustic parameters in whisper-singing vs. whispered speech

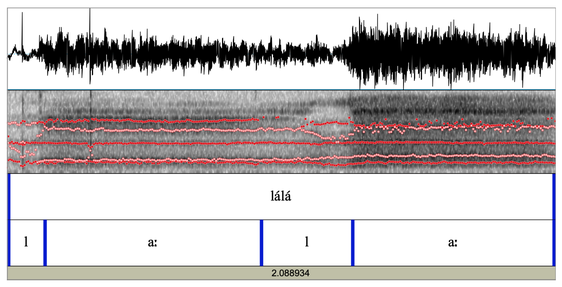
spectral slope → less negative = more effort, higher-pitched melody in whisper-singing vs. speaker's own voice range in whispered speech
F2 → singing may require less precision in vowel quality but more in melody which would enable F2 to shift more in whisper-singing

Read the paper:

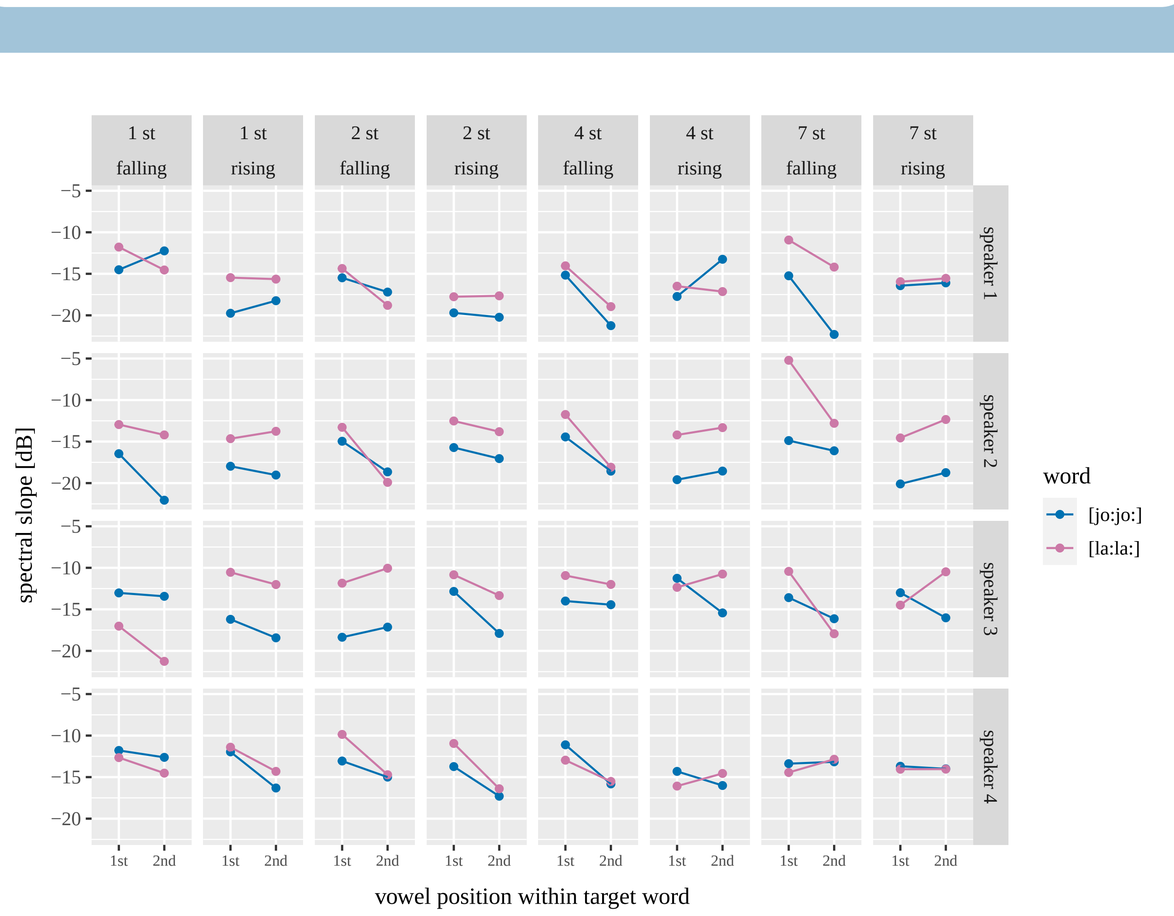
Results

whispered speech ($\alpha = 0.05$)
→ F2 ($p = 0.046$)
→ CoG of signal stop-band filtered between 1000 and 6300 Hz (visible formant structure removed) ($p = 0.022$)
words sung in whisper ($\alpha = 0.05$)
→ F2 ($p < 0.001$)
→ F3 ($p = 0.008$)
→ F2:F1 ($p < 0.001$)
→ F3:F2 ($p = 0.012$)
= movement of F2 more prominent than F1, F3
→ CoG ($p = 0.002$)
→ spectral slope ($p = 0.0069$)

F2 in a perception experiment stimulus recognized in 100% as rising melody:



[la:la:] sung in whisper with a rising melody, intended pitch difference of 7 st. Dotted lines denote formants. Spectrogram frequency range is 0–8 kHz, time step 5 ms.



Spectral slope in the middle third of vowels from words sung in whisper by intended melody, pitch difference and speaker. Each line represents one realization of a target word.

References

1. Meyer-Eppler, W. Realization of Prosodic Features in Whispered Speech. *JASA* **29,** 104–106 (1957).

2. Thomas, I. B. Perceived Pitch of Whispered Vowels. *JASA* **46,** 468–470 (1969).

3. Fónagy, J. Accent et intonation dans la parole chuchotée. *Phonetica* **20,** 177–192 (1969).

4. Higashikawa, M. & Minifie, F. D. Acoustical-perceptual correlates of 'whisper pitch' in synthetically generated vowels. *JSLHR* **42,** 583–591 (1999).

5. Heeren, W. F. L. & Lorenzi, C. Perception of prosody in normal and whispered French. *JASA* **135,** 2026–2040 (2014).

6. Heeren, W. F. L. Vocalic correlates of pitch in whispered versus normal speech. *JASA* **138,** 3800–3810 (2015).

7. Boersma, P. & Weenink, D. *Praat: doing phonetics by computer* (2021). www.praat.org/.

8. R Core Team. *R: A Language and Environment for Statistical Computing* (2022). www.R-project.org/.