Phonetic convergence evaluation based on fundamental frequency variability

Andreeva et al., 2021

Research Questions

1. Do speakers produce different long-term distributional (LTD) pitch profiles depending on the modality of the dialogue and the social status and sex of their interlocutor?

2. Do speakers show convergence and synchrony in mean F0, depending on the modality of the dialogue, and the social status and sex of their interlocutor?

Materials (Speech Corpus)

➤Used subset of Polish Harmonia corpus containing Polish dyadic (i.e., between 2 people) conversations

➤Student-student interactions (**6 female** + **7 male** pairs)

➤Student-teacher interactions (each student w/ same **female university teacher**)

➤Conversation topic: provocative modern art

➤ Scenario 1: Both interlocutors approve of art **[+like; +agree]**

➤ Scenario 2: Both interlocutors disapprove of art **[-like; +agree]**

➤ Scenario 3: Teacher disapproves of art, student accepts **[±like; -agree]**

interlocutors	modality	pair numbers and sex
student-student	[+like; +agree]	6 female, 7 male
student-student	[-like; +agree]	6 female, 7 male
student-teacher	[+like; +agree]	12 female, 14 mixed-sex
student-teacher	[-like; +agree]	12 female, 14 mixed-sex
student-teacher	[±like; -agree]	12 female, 14 mixed-sex

Methodology

➢ F0 (fundamental frequency) taken from inter-pause stretch (IPS) boundaries estimated w/ Robust Algorithm for Pitch Tracking (RAPT) software

➢ Long-term distributional (LTD) measures calculated per IPS:
  ➢ Mean F0 for pitch (Hz)
  ➢ Median F0 for pitch (Hz)
  ➢ Interquartile range (IQR) (Hz)
  ➢ Pitch range (semitones)
  ➢ Standard deviation (SD) for F0 distribution (Hz)

➢ **Convergence** & **synchrony** measured to determine if interlocutors adjust speech when talking to each other (*Edlund et al. (2009)*)

Methodology

➢ **Global convergence:** "similarity between two speakers over the course of a conversation"
  ➢ Measured by feature means over conversation

➢ **Local convergence:** "dynamic similarity between two speakers over time within a conversation"
  ➢ Measured by turn-level similarity

➢ Extracted mean F0 values for first 30% + last 30% of each dialogue

➢ Noted final IPS of first speaker's turn & initial IPS of second speaker (partner)'s turn
  ➢ Mean F0 taken from each turn & normalized by speaker's sex

Findings

➤ **Modality** ([+like; +agree], [-like; +agree], [±like; -agree]) & **sex** (female or male) = most influential factors on mean F0 values

➤ Factors resulting in higher F0 values:

- ➤ Disagreement (i.e., [±like; -agree] condition w/ student accepting & teacher disapproving provocative art)
- ➤ Females = notably higher F0 vs. males
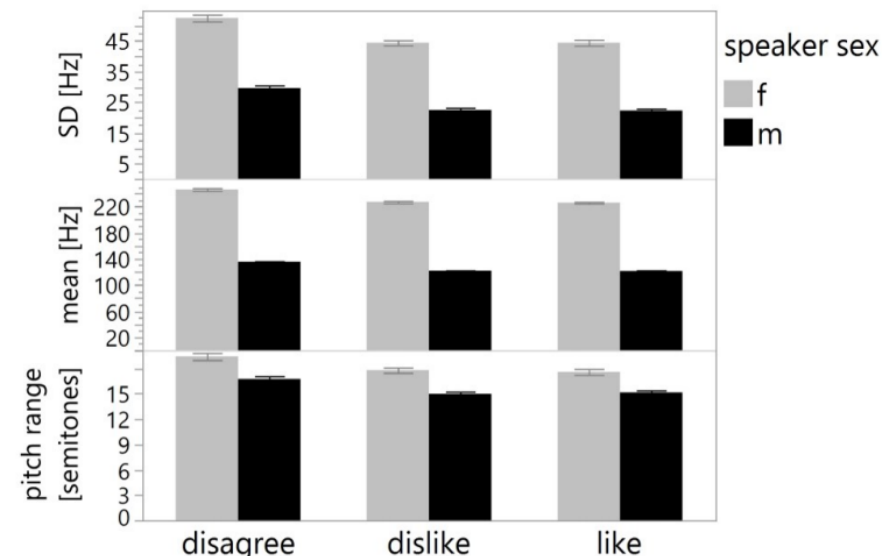- ➤ Speaking to someone in a position of power (teacher)

Table 2 – LTD measures by dialogue modality (standard deviation in parentheses).

parameter	sex	[±like; +agree]	[+like; +agree]	[-like; +agree]
mean (Hz)	m	135.4 (30.1)	121.3 (25.8)	121.8 (23.0)
	f	246.0 (43.5)	225.8 (42.8)	227.1 (46.2)
median (Hz)	m	130.7 (30.2)	118.2 (26.0)	118.6 (22.3)
	f	241.2 (43.5)	222.6 (43.0)	225.1 (45.5)
pitch range (semitones)	m	16.8 (7.8)	15.2 (7.1)	15.0 (7.1)
	f	19.5 (9.2)	17.6 (9.0)	17.8 (8.8)
IQR (Hz)	m	33.0 (31.2)	23.5 (20.9)	24.5 (24.6)
	f	63.5 (40.7)	51.9 (42.6)	53.0 (41.7)
SD (Hz)	m	29.6 (22.0)	22.3 (15.5)	22.5 (18.1)
	f	52.5 (23.8)	44.4 (24.7)	44.4 (22.9)

Most Relevant Info

Consider analyzing different corpora in study

Convergence most prominent in student-teacher interactions & during disagreement conditions; convergence lacking in student-student scenarios

➢Tendency to strive for **agreement** (or lack of disagreement) amongst interlocutors

➢Speakers "converge" to interlocutor **higher** on **social hierarchy** (i.e., <u>student </u>converges to <u>teacher</u>)

Convergence affected by many factors in collaborative tasks, including:

➢Nature of task

➢Hierarchy between participants

➢Perceived attractiveness & likability of interlocutor

➢Visibility of interlocutor

➢Biological sex