Convergence of speaking fundamental frequency in dialogues

Grażyna Demenko, Jolanta Bachan

Institute of Linguistics, Adam Mickiewicz University, Poznań, Poland

lin@amu.edu.pl, jolabachan@gmail.com

Index Terms: phonetic convergence, speaking fundamental frequency, human-human voice communication

1. Introduction

Speaking Fundamental Frequency (SFF), which refers to the pitch level mostly used by the speaker during connected speech, has been considered an important but controversial parameter in voice assessment [1]. Individual adjustments to speech are assumed to subserve the function of controlling social distance. The speaking style of conversational partners thus converges, diverges or remains unchanged, depending on the strategies applied by the interlocutors. The phenomenon of phonetic convergence that occurs naturally and partly automatically in human-human communication has not yet been exploited sufficiently in human-machine communication systems [2]. In principle, each acoustic feature can be the target of convergence processes in production, e.g. voice-onset time, formants, voicing, F0 range and register, intensity, duration, pausing, speaking rate, as well as the long-term average spectrum. This research explores whether the mean F0 of conversational partners would converge over time. The hypothesis is that simple averaging of long term fundamental frequency should help to answer following questions.

1. Does the SFF change during different parts of the dialogue: initial (I), initial-medial (IM), medial-final (MF) and final (F) for the same person and does it depend on the type of dialogue?
2. Does the speaker show any tendency of convergence in their SFF, depending on their speaking role, expressiveness or the aim of the dialogue to be achieved?
3. Is there male/female difference?

2. Methodology

For the corpus, 16 pairs of students (male-male, female-female, female teacher-male students, female teacher-female students) were recorded in a professional studio. In each recording session the pair of speakers conducted the following dialogues: (1) Dialogues NS: Neutral speech, equal role partners, must achieve agreement; (2) Dialogues NNS: Neutral speech, not equal role partners, must achieve agreement; (3) Dialogues ENS: Expressive speech, not equal role partners; (4) Dialogues EES: Highly expressive speech, no role partners; (5) Dialogues EENS: Highly expressive speech, defined role partners. The corpus has been annotated on several layers: orthographic transcription of text, prosody, noise, flow of speaking turns, dialogue acts, agreement and disagreement intervals and speaker's attitude. More details about the corpus can be found in [3]. The dialogues of about 5-min length were divided into 4 sections: I (0-25% of time duration), IM (25-50%), MF (50-75%), and F (75-100%). Each of the sections was about 60-90 seconds long, which provided a relatively stable estimate of the SFF mean values [4].

3. Results

Analysis of variance showed the following results:

1. The SFF changes during different parts of the conversation for the same person.
2. The speaker changes SFF depending on his/her speaking role, type of dialogue, expressiveness.
3. There were statistically significant differences for men and women in dialogues with the teacher, the male speakers lowered their voice, while the female speakers increased their SFF. This is interesting from the point of view of the attractiveness of the conversational partners [5].

4. Discussion

The F0 parameter could be useful in the evaluation of the convergence, but linguistic analysis [6] should also be carried out to interpret in detail the interlocutors' reactions depending on the type of the task, the speakers' role, personality and the gender influence on their behavior. The study has been carried out to create the methodology of the analysis of F0 in measuring the phonetic convergence, standardization of the annotation of pitch changes in dialogue modeling in human-machine communication.

5. Acknowledgements

The study supported by the Polish National Science Centre, project no.: 2014/14/M/HS2/00631, “Automatic analysis of phonetic convergence in speech technology systems”

6. References

[1] Zraick, R. I., Gentry, M. A., Smith-Olinde, L., & Gregg, B. A., “The effect of speaking context on elicitation of habitual pitch”, Journal of Voice, 20(4), 545-554, 2006.

[2] Edlund, J., Gustafson, J., Heldnera, M. and Hjalmarssona, A., “Towards human-like spoken dialogue systems”, Speech Communication 50 (8-9), 630-645, 2008.

[3] Demenko, G and Bachan, J. “Annotation specifications of a dialogue corpus for modelling phonetic convergence in technical systems”, 28th ESSV, 75-82, 2017.

[4] Lindh, J. and Eriksson A., “Robustness of long time measures of fundamental frequency”, 8th Interspeech, 2007.

[5] Hughes, S. M., Farley, S. D., and Rhodes, B. C., “Vocal and physiological changes in response to the physical attractiveness of conversational partners”, Journal of Nonverbal Behavior, 34(3), 155-167, 2010.

[6] Selting, M., “The construction of units in conversational talk”, Language in society, 29(4), 477-517, 2000.