

Investigating the Phonetic Expression of Successful Motivation

Jana Voße^{1,2}, Petra Wagner^{1,3}

¹Phonetics and Phonology Work Group, Bielefeld University, Germany

²Department of Philosophy, Gothenburg University, Sweden

³CITEC, Bielefeld University, Germany

Abstract

The present study provides a comprehensive acoustic phonetic analysis of motivational speech by collecting, annotating and processing 50 minutes of speech data representing less and more successful degrees of motivation. The analysis shows significant differences regarding the acoustic phonetic features f_0 (median, range, variation), intensity (median, range) and speaking rate. We observe inconsistent results for the variation of intensity, pointing to the necessity of a more fine-grained analysis of this feature. This study provides first support for the existence of a specific motivational speaking style.

Key words: acoustic phonetics, motivation, speaking style, emotional speech

Introduction

The concept of motivation is a frequently observed phenomenon in everyday human-human interaction, but also in specific domains like teaching, coaching or nursing. In such interactive situations, linguistic communication is probably the most intuitive way to create a motivational impact. This paper investigates the role of acoustic phonetic parameters within motivational speech.

Although the concept of motivational speech has not been studied intensively so far, we observe research progress on the phonetic expression of related concepts of motivational speech, such as charismatic (Niebuhr et al. 2016) and volitional speech (Skutella et al 2014). These concepts correspond with respect to the characteristics of their acoustic phonetic features, e.g. f_0 , intensity, and speaking rate.

In creating a motivational impact, emotions play a substantial role. Following the concept of emotional empathy, the emotional state of a recipient can be influenced by a speaker's displayed emotion. By expressing a positive emotion, a speaker can set the recipient into a positive state, which in turn influences the recipient's readiness to be motivated positively (Abele 1999). For the expression of emotions, acoustic phonetic features such as speaking rate, f_0 (Burkhardt et al. 2000), and intensity (Tao et al. 2005) are strong means.

Because of the causal relation of phonetics, emotions and motivation and the pragmatic proximity of motivational, charismatic and volitional speech, we hypothesize motivational speech to be characterized similarly. Specifically, we expect motivational speech to be expressed by the following parameters: (1) *Speaking rate*: high number of syllables/second, (2) f_0 : high median (logHz), range (logHz) and variation coefficient, (3) *Intensity*: high median (dB), range (dB) and variation coefficient.

Methodology

We collected, annotated and processed 50 minutes of speech data representing less and more successful degrees of motivation. Based on these, we identified and analyzed our set of acoustic phonetic features potentially relevant for motivational impact. The data consists of the audio extracted from 6 motivational YouTube videos, each presented by a different female speaker aged between 16 and 30 years. The aim of these videos is to motivate their audience to engage in sports and to be on a healthy diet. While presenters' age, gender, video topic and structure as well as upload date are homogeneous, the videos differ in their online ratings. We used these ratings to differentiate between more and less successful motivation. This left us with 3 videos of less successful (15 minutes), and 3 videos of more successful motivation (35 minutes).

The data were force-aligned with AlignTool (Schillingmann et al. 2018) both on a phone and syllable level and corrected manually. Perceptually labeled Interpausal Units (IPUs) are used as a measure of utterance segmentation (mean pause duration = 0.45s). Acoustic phonetic features were measured within IPUs using Praat scripts and served as dependent variables in the subsequent analyses. We assume that they differ significantly between less and more successful levels of motivation. Due to the non normal distribution and high correlation of the dependent variables, statistical analyses are carried out by a series of non-parametric tests (Bonferroni-corrected).

Results

All dependent variables show significant differences between *more motivational speech* (MMS) and *less motivational speech* (LMS), except for intensity coefficient of variation. We further observe higher medians in MMS than in LMS for all parameters, except for intensity coefficient of variation, which shows the opposite case. According to the Brown-Forsythe test, the intensity coefficient of variation, f_0 median and f_0 range show homogeneous variances between MMS and LMS. Regarding the form of

Investigating the Phonetic Expression of Successful Motivation

distribution (tested with Kolmogorov-Smirnov), all dependent variables are characterized by heterogeneous distributions of MMS and LMS.

Table 1. Summary of various test results for all dependent variables.

Dep. variable	MMS (median)	LMS (median)	Wilcoxon rank-sum ('greater')	Brown-Forsythe	Kolmogorov-Smirnov ('two.sided')
Speaking rate (sylls/s)	4.997	4.728	$W = 73383$, $p < 0.01^*$	$F = 30.771$, $p < 0.001^{***}$	$D = 0.1671$, $p < 0.01^{**}$
f_0 median (logHz)	2.385	2.355	$W = 81209$, $p < 0.001^{***}$	$F = 3.9884$, $p > 0.05$	$D = 0.2718$, $p < 0.001^{***}$
f_0 range (logHz)	0.376	0.248	$W = 89347$, $p < 0.001^{***}$	$F = 1.256$, $p > 0.05$	$D = 0.32431$, $p < 0.001^{***}$
f_0 (variation coefficient)	0.705	0.631	$W = 88206$, $p < 0.001^{***}$	$F = 8.7725$, $p < 0.05^*$	$D = 0.30885$, $p < 0.001^{***}$
Intensity median (dB)	73.126	60.147	$W = 114260$, $p < 0.001^{***}$	$F = 158.12$, $p < 0.001^{***}$	$D = 0.65703$, $p < 0.001^{***}$
Intensity range (dB)	37.531	35.851	$W = 77291$, $p < 0.001^{***}$	$F = 6.9092$, $p > 0.05$	$D = 0.15694$, $p < 0.01^{**}$
Intensity (variation coefficient)	0.112	0.136	$W = 37232$, $p > 0.05$	$F = 3.7247$, $p > 0.05$	$D = 0.35986$, $p < 0.001^{***}$

Discussion

We observe statistically significant medians and distributions in MMS and LMS for all dependent variables except for the intensity coefficient of variation (median). Regarding variance, only half the dependent variables show significant results. Obtaining a clear differentiation of MMS and LMS in most dependent variables supports our assumption of a motivating prosodic speaking style contrasting with a less-motivating one. Future perception experiments will investigate whether these production differences are perceptually relevant.

Regarding the assumption of a motivating speaking style, it must be considered that the observed parameter shapes might be speaker-intrinsic rather than articulatorily targeted in a conscious manner, as our study follows a between-subjects design. Testing motivational stimuli in a within-subject design will provide further insight regarding this matter.

Although the results of the analysis of speaking rate, the f_0 parameters, and intensity mean support our hypothesis regarding the relation between successful motivational speech and charismatic, volitional, and positive

emotional speech, we observe differing results regarding intensity variation. A more fine-grained analysis is needed to investigate the role of this parameter further.

For the interpretation, it must be also considered that the chosen unit of analysis (IPU) affects the results. Analysing the given phonetic features on a different level might result in divergent observations.

We are aware that the audio quality of the recorded videos impacts the analysed parameters, especially those of intensity. Hence, the interpretation of the intensity must be considered with reservation. Future experiments with controlled audio qualities will be carried out to examine the validity of the results of the present study. Another point of discussion is the validity of online rankings as a criterion for differentiating levels of more and less successful motivation. Perception experiments are planned to substantiate the approach taken here.

To conclude, our study indicates that successful motivational speech is characterized by a high and variable pitch as well as by a loud and fairly fast articulation, but with a potentially stable intensity within individual utterances.

References

- Abele, A. 1999. Motivationale Mediatoren von Emotionseinflüssen auf die Leistung: Ein vernachlässigtes Forschungsgebiet. In Jerusalem, M. Pekrun, R. (eds.) 1999, *Emotion, Motivation und Leistung*, 31-50. Göttingen, Bern, Toronto, Seattle, Hogrefe-Verlag.
- Burkhardt, F., & Sendlmeier, W. F. 2000. Verification of acoustical correlates of emotional speech using formant-synthesis. In *SpeechEmotion-2000*, 151-156, Newcastle, Northern Ireland, UK.
- Niebuhr, O., Voße, J., & Brem, A. 2016. What makes a charismatic speaker? A computer-based acoustic-prosodic analysis of Steve Jobs tone of voice. *Computers in Human Behavior*, 64 366-382.
- Schillingmann, L., Ernst, J., Keite, V., Wrede, B., Meyer, A. S. & Belke, E. 2018. AlignTool: The automatic temporal alignment of spoken utterances in German, Dutch, and British English for psycholinguistic purposes. *Behaviour Research Methods*, 50(2), 466-489.
- Skutella, L. V., Süssenbach, L., Pitsch, K., & Wagner, P. 2014. The prosody of motivation. First results from an indoor cycling scenario. In Hoffmann, R. (ed.), *Elektronische Sprachsignalverarbeitung 2014*, 71, Dresden, Germany.
- Tao, J., & Kang, Y. 2005. Features importance analysis for emotional speech classification. In Tao, J., Tan, T., Picard, R. W. (eds.), *International Conference on Affective Computing and Intelligent Interaction*, 449-457, Beijing, China.