Interspeech 2022
18-22 September 2022, Incheon, Korea

The Prosody of Cheering in Sport Events

Żygis Marzena[1,2], *Sarah Wesolek*[1,2], *Nina Hosseini*[3], *Manfred Krifka*[1,2]

[1]Leibniz-Centre General Linguistics, Berlin
[2]Humboldt University, Berlin, [3]University of Luxembourg
zygis@leibniz-zas.de, wesolek@leibniz-zas.de, nina.hosseinikivanani@uni.lu, krifka@leibniz-zas.de

Abstract

Motivational speaking usually conveys a highly emotional message and its purpose is to invite action. The goal of this paper is to investigate the prosodic realization of one particular type of cheering, namely inciting cheering for single addressees in sport events (here, long-distance running), using the name of that person.

31 native speakers of German took part in the experiment. They were asked to cheer up an individual marathon runner in a sporting event represented by video by producing his or her name (1-5 syllables long). For reasons of comparison, the participants also produced the same names in isolation and carrier sentences. Our results reveal that speakers use different strategies to meet their motivational communicative goals: while some speakers produced the runners' names by dividing them into syllables, others pronounced the names as quickly as possible putting more emphasis on the first syllable. A few speakers followed a mixed strategy.

Contrary to our expectations, it was not the intensity that mostly contributes to the differences between the different speaking styles (cheering vs. neutral), at least in the methods we were using. Rather, participants employed higher fundamental frequency and longer duration when cheering for marathon runners.

Index Terms: acoustics, speech acts, expressives, emotions

1. Introduction

Cheering is a dramatically understudied communicative activity. [1] developed a classification of speech acts based on performative verbs. The closest to cheering is *to toast*, which expresses a wish, which Austin classifies as a "behabitative" – in particular, an expression "of attitudes to someone else's past conduct or imminent conduct."

In [2], the best-known classification of speech acts, cheers would fit to the class of Expressives that communicate a psychological state. Searle's ([2]) closest example to cheering is 'to *congratulate'*, exemplified by *I congratulate you on winning the race* – but cheering often is done for persons involved in events that are still unfolding, to encourage the person to achieve the goal in the first place. Also, it is not just an expression of the speaker's psychological state but has the goal of supporting and encouraging the addressee (this is a dimension that is neglected in Searle's class of Expressives in general).

One important aspect of cheers is that they are typically collective speech acts – the cheerers are a crowd, and often, as in team sports, the persons cheered to as well. One can plausibly assume that there is more cheering at public sport broadcasts than at broadcasts in private living rooms. One important function of cheering, obviously, is to create a bond between the cheerers. Yet group action is not a defining property of cheers; [3] mentions cheers as one of the many types of speech acts where individual acts can be collectivized.

There are a few observations on the function of cheering in sports. In particular, [4] looks at the "embodied relationship between athletes and cheering", stressing the fact that cheering helps the spectators to join in or participate with the athletes, thus experiencing the same rush of struggle and hope for victory. Without this effect, cheering at sport broadcasts would not make sense at all, as the addressees cannot perceive the cheers. Hence, cheering has an expressive component, similar to exclamatives.

Cheers can be expressed in a multitude of ways – assertions like *You will make it!*, imperatives like *Go on!*, prohibitives like *Don't give up!*, adhortatives like *Let's go!* and non-finite forms like *Faster, faster!* There are dedicated cheering interjections like *hurrah* (even though this is typically used in exclamations). There are idiomatic cheers like established chants of fan clubs, and non-verbal cheers like rhythmical applause, thumbs-up gestures and showing support by dress or mascots. For the linguistic ways of cheering, it appears that the prosodic contour is essential to bring out the cheering quality of an otherwise underspecified speech act.

Cheers are different from calls (summons), as they are not just identifying the referent as addressee (see e.g. [5], [6], [7]). In fact, it would be contra-productive in a race to call out for a runner, as this would probably call him or her down. Cheers are furthermore different from addresses, which identify the social and emotional relationship between speaker and addressee. Addresses like *my dear, honey* or *Professor Smith* that express such relationships cannot easily be used for cheers.

The term *cheering* in its natural use is rather unspecific whether it is used for the communicative support of ongoing events, or congratulations about past achievements, or even consolations about past failures ("cheer someone up"). German makes a rather consistent distinction here: *anfeuern* (lit. 'fire up'), *jubeln* and *aufmuntern* (roughly, 'inciting', 'rejoicing' and 'cheering up').

Here we will investigate the prosody of inciting cheering calls in the first sense, for ongoing events. We expect that there are features of this type of cheering that align with the ongoing event, features that are lacking in the other two types of cheering. For example, applause may tend to be rhythmical and synchronized for the first type, and, at least, initially, non-synchronized for the second.

10.21437/Interspeech.2022-10982

Cheering in the first sense is a case of motivational speech, see also ([8], [9]). The patterns that make speech motivationally successful have not been also studied intensely. In a study of motivational speech on Youtube videos, [10] could show that higher F0 (median, range and variation) leads to a higher degree of successful motivation. At the same time, they observed inconsistent results for the variation of intensity, suggesting a more fine-grained analysis of this parameter.

However, to our knowledge, there are no investigations, about productions of cheering. Thus, the present paper aims at filling in this research gap by investigating how speakers cheer up marathon runners.

2. Experimental evidence

Our research goal is to scrutinize the prosodic realization of inciting cheers for persons that are involved in ongoing competitive events, using the name of that person and how they employ different acoustic parameters including duration, intensity, and F0 to achieve their goals.

To this end, we investigated how the cheering context changes the production of names in contrast to a neutral context. For reasons of comparison, we included names produced in isolation and embedded in sentences in a neutral context. The sentence context was included to avoid a potential confound between sentence and lexical stress and examine whether the names produced in cheering context are more similar to names produced in isolation or sentences. Our research hypotheses were as follows:

Names produced in cheering context are not only louder, but also shorter, and produced with higher F0 than names produced in isolation and in sentences in a neutral context. They are also more similar to names produced in isolation in their acoustic characteristics.

2.1. Experimental design

To test our research hypotheses we conducted an acoustic experiment in which participants produced runners' names while watching a video from a marathon. We also recorded the same names in isolation and embedded in frame sentences.

2.2. Informants

31 native speakers of German took part in the experiments. There were 20 female and 10 female speakers aged 19-50 (mean: 29,7, sd: 10,05). One participant did not fill in the form about the meta data so his/her background data are missing.

2.3. Material

Our material consisted of ten words: five female names and five male names, mostly consisting of sonorants. The number of syllables varied from one to five and the lexical stress appeared on the penultimate syllables apart from one-syllabic words. In Table 1 the stressed syllables are given in bold.

Table 1: *Names used in the experiment.*

Female	Male	Number of syllables
Linn	**Jan**	1
Dag.mar	**Da**.niel	2
Da. **nie**.la	Jo.**han.**nes	3
An.ge.**li**.na	A.le.**xan**.der	4
E.ma.nu**.ʔe**.la	Bar.tho.lo.**mä.**us	5

The recordings were made in three blocks. In the first block the words were read in isolation (three randomized repetitions). In the second block participants read frame sentences with the embedded words (three randomized repetitions). Finally, in the third block participants watched videos with marathon runners and their task was to produce the runner's name while watching the video form a marathon (see section 2.4. for the procedure)

We ended up with the recordings of 12900 items that were submitted to further analysis.

2.4. Procedure

The experiment took part in a sound-proof lab at the Leibniz-ZAS in Berlin. Block 1 (items in isolation) and block 2 (items embedded in carrier sentences) were presented in Power Point and block 3 (the videos) in the VCL media player. The participants were recorded by using a microphone (Sennheiser MKH 20 P48 with a sampling rate of 48 kHz, 16 bit) that was positioned ca. 20 cm from the speaker.

Before starting the experiment, participants were informed about the order of presentation of the three blocks. They were also asked to imagine that one of their friends is running in a marathon and they should cheer for him or her up by using his or her name. The runner was marked (overshadowed) with a moving light on the video so that it was easy to follow him/her during the run. In the left bottom corner of the video, the name of the runner was displayed, see Figure 1. There were a series of videos. In each video, either a female or male runner was shown, again presented in the moving light and with his/her name. The time of the videos varied from 0.28-0.37s.

The participants were assured that they would not be heard by anyone so that they can screen and behave like real fans. After the experiment was over, several participants said that they enjoyed the experiment because it was different from other experiments they had participated in and they were able to "goof off". Several people also said it was to some extent therapeutic for them.

Figure 1: Screenshot *of a video used in the experiment for the stimulus "Dagmar".*

2.5. Measurements

The data were annotated and analysed in PRAAT (version 6.0.40 [11]).

After a semi-automatic annotation, the segment boundaries were manually checked by one annotator and then by another annotator.

The following parameters have been measured:

- duration of the name and all its syllables

- maximum and mean intensity of the word and all its syllables
- F0 maximum and F0 mean of the word and all its syllables

In addition, we also calculated speech rate by dividing the number of syllables of a given word by the duration of the word for all contexts (items in isolation, embedded in frame sentences, and cheering items).

2.6. Statistics

The statistical analysis was conducted in R 3.4.2 ([12]) by using packages lme4 (Bates et al. 2018, [13]) and *emmeans* ([14]).

We built linear mixed models with *Duration, Intensity*, and *F0* as dependent variables, as described in section 2.5 and Context Type [items in isolation, items embedded in frames, cheering items], the Number of Syllables of names [1-5], participant's Sex [male, female], grammatical Gender of Names [male, female], Speech rate and participant's Age as fixed effects. We included the interactions of participant's Sex and the grammatical Gender of names. We also added Item and Participant as random intercepts, as well as random by-participant slopes for the Context Type, the Number of syllables, and Speech rate, and a random by- item slope for the Style type and Speech rate. Due to convergence issues, some of the random slopes had to be removed. For the duration, we dispensed with Speech rate as a factor and random slope because Speech rate was calculated by including word duration (i.e the number of syllables was divided by word duration) and the two measurements, i.e. Speech rate and duration, were confounded to some extent.

We started with full models, removed non-significant effects, and selected the best fit by means of likelihood ratio tests with the ANOVA function in R. Since the variables Style Type and Number of syllables consisted of more than two levels we corrected for multiple tests by using pairwise comparisons available in the emmeans package.

2.7. Results

Our results show that names produced to cheer up runners are mostly produced as sequences of names repeated with similar intensity, duration and F0 patterns, see Figure 2.

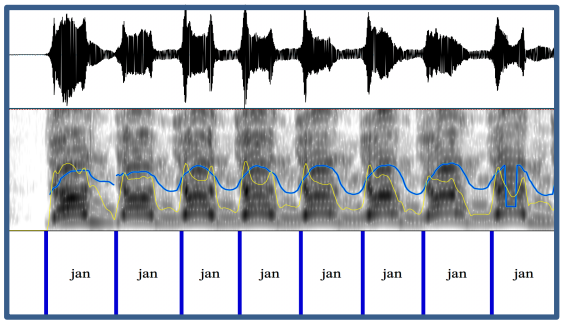

Figure 2: *A spectrogram and oscillogram of the name "Jan" produced in a cheering context*

Several participants also produced the multi-syllables names by dividing them into syllables with similar duration. Others produced them as quickly as possible without any perceptible division into syllables. They often put more emphasis on the first syllable in names that were three and more syllables long. That is, lexical stress on the penultima was shifted on the initial syllable. Finally, it should also be mentioned that a few speakers also employed both strategies.

The results reveal that cheering items were longer in comparison to items spoken in isolation (t= 6.448, p<.001) and to items embedded in frame sentences (t=9.149, p<.01), see Figure 3. As expected, the duration was also dependent on the number of syllables (p<.001 for all comparisons), see Figure 3.

It is also worth mentioning that the speech rate was lowest in the cheering context (mean 3.34, sd. 1.31), followed by the isolated context (mean 4.48, sd. 1.37) and by sentence context (mean 5.81, sd.=1.65).

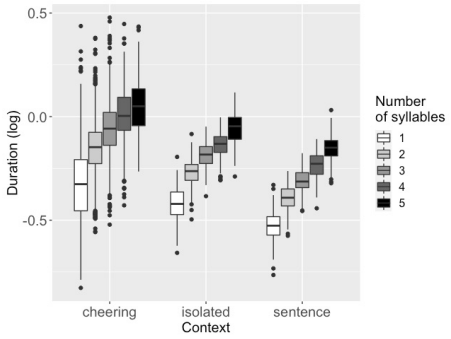

Figure 3: *Duration of items across contexts*

Regarding intensity, it was higher in the cheering context than in isolated and sentence context, but the difference did not reach significance. However, it turned out that longer words were pronounced with a higher max intensity (t=2.40, p<.05) and increasing speech rate also increased the intensity (t=2.50, p<.05).

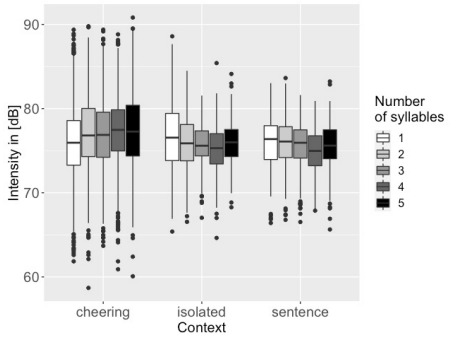

Figure 4: *Max intensity of items across contexts*

Male speakers were louder than female ones (t=2.95, p<.01). They also pronounced female names with more intensity than male ones while female participants pronounced male names with more intensity than female ones. The interaction of participant's sex and the grammatical gender of names was significant (t=-3.50, p<.001).

With respect to mean intensity, our results show that names produced in a cheering context were louder as compared to names produced in embedded sentences (t=-15.42 p<.001), and also names in isolation were produced with greater intensity as compared to names embedded in sentences (t=12.69, p <.001). However, no significant difference was found between the intensity of items produced in cheering vs. isolated context.

The results were contrary to our hypothesis concerning the intensity and somewhat contra-intuitive. However, a closer inspection of individual speakers revealed that twelve speakers

produced the highest max intensity in the cheering context as opposed to isolated and sentence context (vp:1, 2, 3, 5, 16, 19, 23, 24, 25, 27, 39).

The F0 peak was highest in names in the cheering context as compared to names in the isolated context (t=10.96, p<.001) and embedded sentences (t=56.46, p<.001). Figure 5 presents the results. As expected, female speakers produced the names with a higher peak than male speakers did (t= -2.55, p<.05). Finally, female speakers produced female names with a higher F0 peak than male names, and conversely, male speakers produced male names with a lower F0 peak than female names, but the difference did not reach statistical significance (t= 1.86, p=.06).

Very similar results were obtained with respect to the mean F0 with an additional significant effect of the number of syllables: longer words showed a lower F0 mean (t=-2.19, p<.05).

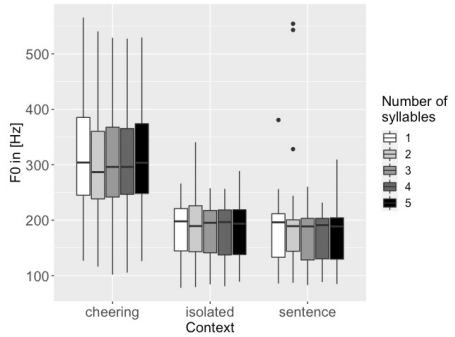

Figure 5: *Max F0 as produced across contexts*

3. Discussion

Our data revealed that names produced as cheering items are different from those produced in isolation and in embedded sentences produced in neutral speech mode.

We expected a greater intensity in the cheering situation. The rationale for this is that cheerers experience greater arousal than in the comparison situations, and/or iconically express greater support with the exertion of more energy. However, contrary to our expectations, it was not the intensity that contributed most to the differences between the cheering and the comparison situations. Rather, it was the fundamental frequency and duration that are employed by speakers to cheer up marathon runners.

However, regarding intensity, it should be stressed that about 1/3 of our speakers produced higher intensity in the cheering context which is in line with our hypothesis. As for other speakers, it might be the case that since the intensity is a more sensitive parameter requiring strict distance controlling and our speakers could have unconsciously changed it, the resulting differences or their lack of could have been caused by distance changes. Furthermore, even if our speakers might have felt relaxed the experiment still took place in laboratory settings that might have imposed limitations to at least some of our participants. Also, the increasing intensity is expected when the distance from the speaker increases or when someone is far away. This was not the case in our lab when the results were so much dependent on the speakers' imagination and their willingness to adapt themselves to a potentially real situation.

In our results, higher F0 is the most powerful parameter employed by speakers to produce cheers. This is in line with findings about higher F0 means in the expression of high arousal emotions, see ([15]). Higher F0 can be seen as expressing higher arousal of the speaker but also as the intention to instill higher arousal in the addressee, thus supporting the addressee in his or her efforts.

[10] have also found that higher F0 (median, range, and variation) leads to a higher degree of successful motivation. At the same time, they observed inconsistent results for the variation of intensity, suggesting a more fine-grained analysis of this parameter. Similarly, [8] reported that a higher F0 level (and range) was used in more charismatic speeches by Steve Jobs as compared to less charismatic speeches by Mark Zuckerberg but the intensity mean was found to be significantly higher in the investor-oriented speeches held by Steve Jobs; no intensity difference was found in the customer-oriented speeches between the speakers.

We also found differences in speech rate: Names in the cheering context were produced with the slowest speech rate as compared to names produced in other contexts. We did not expect this, as the speakers aimed to motivate the runners to speed up, so one would expect the rate of the motivating speech to be increased as a matter of iconicity. However, greater duration may also convey iconically greater stamina. Closer inspection of the data revealed that speakers very often prolonged the duration of selected syllables, or they cheered up the runners by syllabifying the names and therefore introducing short pauses between the syllables of a given name. This is reflected in the duration measurements: the cheering items were significantly longer and showed more variation than those produced in isolation and sentences. This finding leads us to reject our initial hypothesis about the shortest duration of cheering items. A future study will explore which syllables were more prolonged in comparison to others and to what extent the lexical accent did play a role here.

4. Conclusions

Our study of cheering contours reveals that speakers use different strategies to meet their communicative goals, i.e., to motivate marathon runners. Some speakers produced the runners' names by dividing them into syllables while others pronounced them as quickly as possible putting the emphasis on the first syllable of a multisyllabic name. All speakers repeated their cheers by uttering the name several times. The motivation for this is evidently to express continued support for an ongoing event. Also, repetition may convey greater stamina to the addressee. In our study, the repetitions were nearly always rhythmically spaced out in equal intervals. This may support the rhythmical movements of the athlete, by aligning the speech signal to these movements, and it might also iconically convey stamina and the determination of not giving up.

In summary, our results show that speakers employed higher F0 and longer duration when cheering for marathon runners. Results of intensity were rather unclear requiring further investigation.

5. Acknowledgments

Research on this paper was done in the context of the ERC Advanced Grant SPAGAD: Speech Acts in Grammar and Discourse, receiving funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 787929. We would like to thank all participants for taking part in our experiment.

6. References

[1] J. L. Austin, *How to do things with words*. Oxford: Oxford University Press, 1962.

[2] J. R. Searle, *Speech acts: An essay into the philosophy of language*. *Cambridge University Press, 1969*.

[3] K. Ludwig, "What are group speech acts?" *Language and Communication,* vol. 70, pp. 46-58. 2020.

[4] K. Bicknell, "Sport, entertainment and the live(d) experience of cheering". Popular Entertainment Studies, vol. 2, 96-111. 2011.

[5] J. Pierrehumbert, The Phonology and Phonetics of English Intonation. Ph.D thesis, MIT. 1980.

[6] D. R. Ladd, *Intonational Phonology*. Cambridge University Press, Cambridge. 2008.

[7] A. Arvaniti, M. Żygis, and M. Jaskula, "The phonetics and phonology of the Polish calling melodies". *Special Issue of Phonetica*, vol. 73, pp. 338–361, 2016.

[8] O. Niebuhr, A. Brem, J. Michalsky, and J. Neitsch, "What makes business speakers sound charismatic? A contrastive acoustic-melodic analysis of Steve Jobs and Mark Zuckerberg". *Cadernos de Linguistica e Teoria da Literatura*, vol. 1, no. 1. 2020.

[9] L. V. Skutella, L. Süssenbach, K. Pitsch, and P. Wagner, "The prosody of motivation. First results from an indoor cycling scenario". In Hoffmann, R. (ed.), Elektronische Sprachsignalverarbeitung 2014, 71, Dresden, Germany.

[10] J. Voße and P. Wagner, "Investigating the Phonetic Expression of Successful Motivation". *ExLing 2018*, Paris, France, 2018.

[11] Boersma, P., and D. Weenink (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.40, retrieved 5 July 2018 from http://www.praat.org/

[12] R Core Team (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

[13] Bates, D., Maechler, M., Bolker, B. and S. Walker (2018). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–17.

[14] Lenth, R (2019). emmeans: Estimated Marginal Means, aka Least-Squares Means.

[15] T. Bänziger and K. R. Scherer, "The role of intonation in emotional expressions", *Speech Communication,* vol. 46, pp. 252-267, 2005.