SOME ASPECTS OF PROSODY OF FRIENDLY FORMAL AND FRIENDLY INFORMAL SPEAKING STYLES

Dmitry Sityaev, Gabriel Webster, Norbert Braunschweiler, Sabine Buchholz, Kate Knill

Speech Technology Group, Toshiba Research Europe Limited, Cambridge, UK
{dmitry.sityaev,gabriel.webster,norbert.braunschweiler}@crl.toshiba.co.uk

ABSTRACT

The current study investigates acoustic correlates associated with friendly formal and friendly informal speaking styles. A small corpus of speech was recorded by a native speaker of American English. The results revealed that that the most distinctive feature differentiating the two styles is the fundamental frequency. There was also a small difference found in the articulation rate and RMS energy between the two styles.

Keywords: speaking styles, prosody.

1. INTRODUCTION

Expressive speech has become a very popular subject of research nowadays amongst phoneticians, psychologists, engineers working in speech technology and so on. A significant amount of work has already appeared on expression and recognition of emotions, and cross-language and cross-culture studies have also shown the universal nature of emotion. Emotional speech synthesis is also one of the hottest topics in the area of speech industry (for a good overview, see [7]).

Speaking styles have also been recently treated under the rubric of expressive speech. Research into speaking styles has not been so widespread as research into emotions yet, however, its popularity is slowly gaining weight. One of the major problems is there are no clear definitions of speaking styles in the literature. Researchers either structure their studies around dichotomies such as spontaneous vs. read speech or slow vs. fast speech, or they pick up arbitrary speaking styles and compare them (e.g. literary novel style vs. advertising style vs. encyclopedia speaking style, [1]).

An ESCA workshop "The Phonetics and Phonology of Speaking styles" which took place in 1991 in Barcelona, Spain saw a good body of work produced on the subject. It also demonstrated that there is no standardisation of labels and definitions of speaking styles which does not help to bridge the work carried in the field of phonetics with the work carried by sociolinguists.

This paper presents the results of acoustic study of friendly formal and friendly informal speaking styles (referred to from here onwards as "formal" and "informal") obtained from one American English speaker. Despite claims in [3] that actors' speech is not always optimal for the study of emotions/speaking styles, we used a professional voice talent to record our prompts. This was done simply for the reason that we had previously recorded a neutral style using the same speaker. Two methods are proposed in [5] which can be used for eliciting expressive speech using actors – elements from both of these were used in the recording of the corpus. We also conducted a perceptual cross-validation study to assure that an intended speaking style is recognised as such and discarded samples that were not perceived as intended.

Our hypotheses were that the informal style is characterised by a wider pitch range and faster rate of articulation. We also expected to have a higher average number of intermediate phrases per intonational phrase in the informal style compared to the formal style – a finding which is reported in [4].

Section 2 describes the experimental setup. Section 3 provides analysis of the results. Discussion of the results is provided in Section 4, while Section 5 serves as a conclusion.

2. METHOD

2.1. Corpus design and recording

A corpus of 625 sentences was designed to elicit formal and informal styles. Consideration was given to the following: all sentences must be easily imaginable to be uttered in formal and informal situations. It was also ensured that the corpus contained both short and long sentences as well as provided a good phonetic coverage.

The voice talent was presented with a question/answer pair and was asked to only read the question part to herself, imagine the situation, and utter the answer part. First, 625 sentences were recoded in the formal style, where the voice talent was instructed to speak as if she was talking to a work colleague or a customer. Then, the same 625 sentences were recorded in the informal style, where the voice talent was instructed to speak as if she was talking to a close friend. In both instances, the voice talent was asked to sound friendly.

A total of 1250 sentences were recorded. To enable us process pitch information more accurately, Laryngograph signal was also captured.

2.2. Data cross-validation

Following the recording of the corpus, a perceptual experiment was designed with a purpose of data cross validation. Since the total number of sentences appeared to be large to be presented to each listener, only 200 sentences were selected from each corpus (i.e. 200 sentences from the formal corpus and 200 sentences from informal corpus). Additionally, 200 sentences were chosen from the neutral corpus available for the same voice talent – these were intended as distractors. All sentences were different to avoid repetitions.

For each utterance played, listeners were asked the question: "Does the speaker come across as sounding formal or informal?" The options given to the subjects as an answer were: "formal (as if talking to a client)", "informal (as if talking to a friend)" and "neither". The subjects could listen to any sentence as many times as they wanted to.

A total of 10 subjects took part in the experiment. All subjects were native speaker of American English.

2.3. Analysis

The results of the data cross validation experiment are presented in Table 1 below. It appears that the formal style was also often perceived as informal.

Style	formal	informal	neither
Formal style	44%	42%	14%
Informal style	23%	71%	6%
Neutral style	85%	8%	7%

Table 1: Perception of the three styles as either formal, informal or neither for all presentations by all subjects.

All data were automatically segmented and annotated. Pitch tracks were obtained from the Laryngograph signal and used in hand labelling the data with ToBI mark-up.

Acoustic analysis was carried only on those utterances where the degree of formality was recognised "correctly" (as intended) by 6 a more people (i.e. above chance). That comprised 77 sentences for the formal speaking style and 145 sentences for the informal speaking style.

A number of acoustic parameters were investigated. Since the two subsets analysed do not happen to be same utterances, the analysis was more or less restricted to the analysis of global parameters (e.g. mean F0, articulation rate, etc.). In all cases, Welch two-sample t-tests were carried out.

3. RESULTS

3.1. Fundamental frequency

The mean F0 (standard deviation) for formal sentences is 194Hz (43) and the mean (standard deviation) for informal sentences is 247Hz (80) – see Figure 1 below. The mean differences for F0 were found to be statistically significant ($t=-23.99$, $df=212$, $p<0.001$).

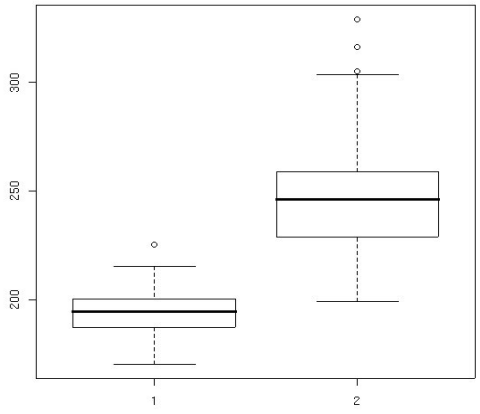

Figure 1: Mean fundamental frequency for formal and informal speaking styles.

The sentences belonging to the informal style were also characterised by a bigger F0 range compared to the sentences recorded in the formal style – the average range value for the informal set is 386 Hz and the average range for the formal set is 236 Hz. Again, the differences were found to be statistically significant ($t=13.32$, $df=140$, $p<0.001$).

T-tests revealed that F0 utterance initial values were different between the two styles: 190Hz for the formal style and 223Hz for the informal style. The difference was statistically significant (t=-6.37, df=211, p<0.001). F0 utterance final values were also different: 131Hz for the formal style and 172Hz for the informal style. Again, t-test showed that the difference was statistically significant too (t=-5.15, df=172, p<0.001).

3.2. RMS energy

RMS energy was found to be significantly affected by the style (t=-3.60, df=136, p<.001). The RMS energy for the formal style (50.7 dB) was a bit lower than the RMS energy for informal style (52.1 dB).

3.3. Articulation rate

Articulation rate was calculated as the number of segments per second. A small difference was found between the formal and informal sentences – 11 vs 10 segments per second, however, the significance was quite small (t=2.27, df=177, p=0.02). Thus, formal style appears to be characterised by a slightly faster speech.

3.4. Pause duration

The duration of sentence-internal pauses was also compared for the two styles. There was no statistically significant difference in sentence-internal pause durations between the two styles (t=-1.67, df=127, p>0.05).

3.5. Pitch accent distribution

Table 2 below demonstrates the pitch accent distribution for the formal and informal styles.

Pitch accent	Formal	Informal
H*	38%	57%
L+H*	44%	26%
!H*	8%	9%
L*	1%	4%
L+!H*	7%	3%
Other	2%	1%

Table 2: Percentage of pitch accent by pitch accent type for the formal and informal styles.

In general, the majority of the utterances (82-83%) in both styles contained H* and L+H* pitch accents. However, there were more H* accents than L+H* accents in the informal style, whereas in the formal style, they were distributed a bit more equally.

3.6. Phrase accent and boundary tones

All utterances were also examined with respect to F0 behaviour utterance-finally. Out of 77 sentences in the formal speaking style, 76 sentences were characterised by a falling F0 (L-L%) and 1 sentence was characterised by rising F0 at the end of the sentence (H-H%). The majority of the informal sentences (84%) were also characterised by a falling F0 utterance-finally (L-L%), however, there were more sentences with a rising patterns (L-H%, H-H%) compared to the formal style – 15%. One sentence was finished with a level tone in the informal style.

3.7. Prosodic phrasing

For each sentence, we calculated the average number of intermediate phrases per intonational phrase. There was no difference in the average number of intermediate phrases per intonational phrase between the two styles (t=1.26, df=139, p>0.05), the averages being 1.9 for the formal style and 1.7 for the informal style.

4. DISCUSSION

The results of the acoustic analysis revealed that the formal and informal speaking styles do differ in some acoustic parameters. Since the utterances selected were not parallel, the analysis was primarily limited to global variables, i.e. F0 mean, F0 range, etc.

Fundamental frequency was by far the most distinctive correlate as far as the formal and the informal styles are concerned. Not only were the sentences belonging to the informal style characterised by higher average F0, their pitch range was also higher than the pitch range of the sentences belonging to the formal style. Gussenhoven (2004) mentions that amongst various interpretations of the Frequency Code for higher pitch is “friendliness”. Although both styles were meant to be “friendly”, this affective interpretation of the Frequency Code probably surfaced more in the friendly informal style and thus explains the difference observed between the informal and the formal speaking styles.

Additionally, the utterance-initial F0 values and utterance-final F0 values were higher in the informal style than in the formal style. This difference is very likely to be caused by the change in the mean F0 used. The informal style also contained a higher proportion of sentence with a rising F0 pattern utterance-finally. Again, final rises (higher pitch) may be another use of the Frequency Code to convey a higher degree of "friendliness", and ultimately "informality".

Our results did not support findings mentioned in [3] with respect to prosodic phrasing – there was no difference found in the number of minor tone units (i.e. intermediate phrase) per major tone unit (i.e. intonational phrase).

Contrary to our hypothesis set above, the formal speech was characterised by a slightly faster articulation rate (i.e. shorter durations) than the informal speech. It has been shown that faster speakers come across as more convincing and more confident [2] which is a sign of competence – a feature often associated with the formal speech. Also, in the informal speaking style, a speaker may do some cognitive processing on the fly, which effectively, introduces more variation in the speaking rate, thus possibly slowing it down overall.

There was also a small difference in the RMS energy between the formal and the informal utterances. The RMS energy was slightly higher for the informal style, however, it is questionable whether such a small difference is perceptually significant.

Formal and informal styles were characterised by more or less similar distribution of pitch accents. However, they was a high proportion of H* accents in the informal style. It is not clear whether this is due to the style itself or a result of the grammatical structure of the text.

5. CONCLUSION

The comparison of the friendly formal and friendly informal styles revealed that the fundamental frequency appears to be a most prominent feature distinguishing the two styles. The friendly informal style sentences have higher average F0 and higher F0 range than the friendly formal style sentences. The friendly informal style was characterised by a slightly higher RMS energy and a slightly slower articulation rate. Although a voice talent was used to record stimuli, data cross validation experiment showed the informal style was perceived as such more often compared to the formal style which was sometimes perceived as formal and sometimes as informal.

It should be borne in mind that the findings of this study are based on the analysis of the performance of one speaker. More experiments are needed to establish whether other speakers use similar or different acoustic cues when speaking in the friendly formal or friendly informal styles.

6. REFERENCES

[1] Abe, M. 1991. Speaking styles: statistical analysis and synthesis by a text-to-speech system. In van Santen, J., Sproat, R., Olive, J., Hirschberg, J. (eds), *Progress in Speech Synthesis*. New York: Springer-Verlag, 495-510.

[2] Apple, W., Krauss., R. 1979. Effects of pitch and speech rate on personal attributions. *Journal of Applied Social Psychology* 37, 715-727.

[3] Campbell, N. 2000. Databases of emotional speech. *Proceedings of the ISCA Workshop on Speech and Emotion*, Belfast, 34-38.

[4] Cid, M., Fernandez Corugedo, S.G. 1991. The construction of a corpus of spoken Spanish: phonetic and phonological parameters. *Proceedings of the ESCA Workshop "Phonetics and Phonology of Speaking Styles: Reduction and Elaboration in Speech Communication"* Barcelona, 17-1 – 17-5.

[5] Enos, F., Hirschberg, J. 2006. A framework for eliciting emotional speech: capitalising on the actor's process. *Proceedings of the Satellite Workshop "Corpora for research on emotion and affect"*, Genova, 6-10.

[6] Gussenhoven, C. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

[7] Schröder, M. 2001. Emotional speech synthesis: a review. *Proceedings of Eurospeech*, Aalborg, 561-564.