Acoustic analysis of politeness and efficiency in a cooperative time-sensitive task

Marcela Charfuelan[1], Paul M. Brunet[2]

[1] DFKI GmbH, Language Technology Lab
Alt-Moabit 91c, D-10559, Berlin, Germany
{marcela.charfuelan}@dfki.de
[2] School of Psychology, Queen's University Belfast, United Kingdom
{p.brunet}@qub.ac.uk

Abstract. We present an acoustic analysis of politeness and efficiency in a cooperative time-sensitive task experiment. In the experiment sixteen dyads completed 20 trials of the "Maze Task", where one participant (the *navigator*) gave oral instructions for the other (the *pilot*) to follow. For half of the trials, navigators were instructed to be polite, and for the other half to be efficient. We investigate what are the main acoustic factors that are associated with greater politeness in the polite condition and lesser politeness in the efficient condition.

1 Introduction

Detection, analysis and synthesis of social signals are topics increasingly applied in computing technologies. Sensitive Artificial Listeners (SAL), which are machines that possess some social and emotional intelligence capabilities [7], pedagogical agents that exhibits social intelligence [10] or predictors of behavioural outcomes in social situations [8] are just some examples where social signals play and important role.

Social signals like politeness, empathy, hostility, (dis)-agreement and any other stances towards others, can be expressed through verbal and non-verbal means in different modalities [9]. One of these modalities is *vocal nonverbal behaviour* – not *what* is said, but *how* it is said. This includes prosodic features such as pitch, energy and rhythm, as well as voice qualities such as harsh, creaky, tense, etc. Regarding politeness, Brown and Levinson [1] predicted that sustained high pitch (maintained over a number of utterances) will be a feature of negative-politeness usage, and creaky voice a feature of positive-politeness usage, and that a reversal of these associations will not occur in any culture.

Social signals, like politeness, typically occur in interactions among people; this makes it natural to study them in corpora of spontaneous interactions rather than in material produced by an actor out of context [4]. In this study we analyse the recordings of a cooperative time-sensitive task experiment designed to study vocal expression of politeness and efficiency [2]. In the experiment sixteen dyads completed 20 trials of the "Maze Task", where one participant (the *navigator*)

gave oral instructions (mainly "up", "down", "left", "right") for the other (the *pilot*) to follow. For half of the trials, navigators were instructed to be polite, and for the other half to be efficient. In this experiment, task accuracy is an objective measure calculated by the distance from the cursor position at the end of the trial and the end point.

In a preliminary analysis of the experiment, it was found that although the task was very simple and users had few ways to express politeness, it significantly affected task accuracy and pilots's subjective ratings indicate that it was perceived [2]. So in this paper we investigate what are the main acoustic factors that are associated with greater politeness in the polite condition and lesser politeness in the efficient condition. We use Principal Component Analysis (PCA) to analyse possible clusters on the data and multiple linear regression to find the acoustic features that better predict task accuracy. If the task accuracy is systematically affected by the politeness/efficiency condition we would like to know whether there are predominant acoustic features in each condition.

The paper is organised as follows. In Section 2 we start describing the experiment, data and methodology used in this study. Then in Section 3 we briefly describe the acoustic measures extracted from the data. Results are presented in Section 4 and conclusions in Section 5.

2 Data and method

The study consisted of participants engaging in a cooperative task with a partner. The participant was positioned in front of a computer monitor in one room, while the partner was in a second room. The assigned task was a computerized maze task requiring the dyad to guide the cursor from the starting point of the maze to the endpoint. The participant could see the maze on the computer monitor but did not have the means to directly move the cursor. The other dyad member could not see the maze (instead they saw the participants face via a webcam) but with the arrow keys of the keyboard could move the cursor. The dyad could communicate via microphones and speakers. Consequently, the participant had the role of navigator and was responsible for verbally guiding the partner's cursor movements. In total the dyad completed 20 trials. The experimental trials were broken into 4 blocks of 5 trials. In each block, the trials became increasingly more difficult by increasing the black squares by 5%, also for the second and fourth blocks the vertical and horizontal cursor controls were flipped (participants were informed of this change). For the first 10 trials, the participant was instructed to be polite, the second 10 trials to be efficient. Half of the participants were instructed to be efficient first, then polite for the second part. The trials were time sensitive (less than a minute allotted) and errors (i.e. hitting the walls) decreased the allotted time limit.

The blocks and trials of every session and the words or command words used by the navigators were manually segmented. Acoustic features were extracted from these small segments and averaged if the extracted measure is frame based. The distribution of data is presented in Table 1, due to technical problems with

the recordings we have analysed 14 of the 16 dyads, corresponding to 4 male and 10 female navigators. In this table the data has been split according to the difference score (Diff score) between the average accuracy scores of the polite and efficient sessions.

Table 1: Distribution of data. Diff score is the difference between the task accuracy score obtained on the polite condition and the score obtained on the efficient condition.

	Diff score > 10		Diff score ≤ 10		
Condition	female	male	female	male	Total
efficient	1127	379	958	562	3026
polite	1452	517	959	382	3310

For the analysis of the data, first we use Principal Component Analysis (PCA) to analyse possible clusters on the data and the two conditions. Then we perform multiple linear regression using the task accuracy score of each trial as objective measure and several acoustic features as explanatory variables. We search for the acoustic features that better predict the accuracy score of each trial using ten repetitions of ten-fold sequential floating forward selection - multiple linear regression (SFFS-LM).

3 Acoustic measures

The acoustic measures used in this study are described in detail in [3], here we mention them briefly:

1. Low level acoustic measures
   - Voicing strengths: full-band and multi-band: str, str1, str2, str3, str4, str5
   - Pitch harmonics magnitude: first ten magnitudes: mag1...mag10.
   - Spectral features: Melcepstrum coefficients (mcep0...mcep24), Spectral entropy (full-band and multi-band: spec_entropy, spec_entropy1,..., spec_entropy5)
   - Articulatory-based features: Formants, Formant bandwidths, Formant dispersion
2. Prosody acoustic measures
   - Fundamental frequency or pitch
   - Pitch entropy (calculated as the spectral entropy)
   - maximum, minimum, and range of f0
   - Duration of the utterance in seconds
   - Voicing rate calculated as the number of voiced frames per time unit
   - Energy, calculated as the short term energy $\sum x^2$
3. Voice quality acoustic measures

- Hamm_effort = LTAS_{2-5k}
- Hamm_breathy = (LTAS_{0-2k} - LTAS_{2-5k}) - (LTAS_{2-5k} - LTAS_{5-8k})
- Hamm_head = (LTAS_{0-2k} - LTAS_{5-8k})
- Hamm_coarse = (LTAS_{0-2k} - LTAS_{2-5k})
- Hamm_unstable = (LTAS_{2-5k} - LTAS_{5-8k})
- slope_ltas: least squared line fit of LTAS in the log-frequency domain (dB/oct)
- slope_ltas1kz: least squared line fit of LTAS above 1 kHz in the log-frequency domain (dB/oct)
- slope_spectrum1kz: least squared line fit of spectrum above 1 kHz (dB/oct)

Low level acoustic measures are extracted at frame level, with a frame length of 25 ms. and a frame shift of 5 ms. The frame based measures are averaged per word. Prosody features are classical features related to pitch, energy, duration, etc. And voice quality measures are measures mostly used in emotion research. Prosody and voice quality measures are extracted at word level.

4 Results

4.1 PCA analysis

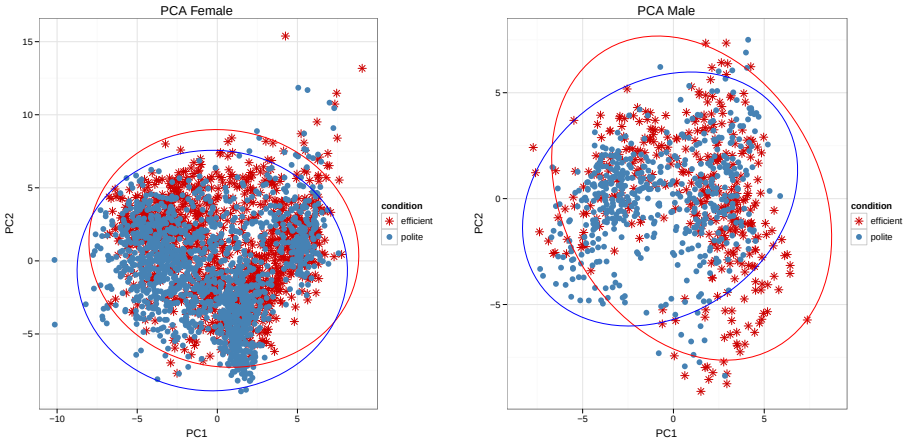

Fig. 1: PCA analysis of male and female data with Diff score > 10. The first two PCs in female data explain 32% of the variance and in male data the first two PCs explain 24% of the variance.

Since we do not have perceptual annotations of how polite the users were when they were asked to be polite, just their subjective impressions collected through a questionnaire, for the first experiment with PCA we selected the

sessions where the difference score is high. That is, the sessions where the task accuracy score obtained on the polite condition was higher than the score on the efficient condition (in this experiment a high score means low accuracy). As we mentioned in the introduction, in a preliminary study it was already detected a consistent acoustic separation in individual sessions where the polite and efficient scores were very different.

In Figure 1 a scatter plot of the first two principal components of the PCA analysis is presented. In this analysis we have used all the acoustic features and the data where the Diff score is > 10 (see Table 1). We expected that the clusters were more apparent when there is a big score difference between the polite and efficient condition. An ellipse in these figures indicate clusters of words used used during the polite and efficient sessions. The clusters for male data seem to be more separated than for female data, but there is also less male speakers in this data. PC1 in both cases separate better the clusters.

Table 2: Main loadings for acoustic features for the male and female PCA analysis presented in Figure 1.

Female PCA				Male PCA			
PC1		PC2		PC1		PC2	
Feature	score	feature	score	Feature	score	feature	score
spec_entropy1	-0.22	spec_entropy4	-0.23	mcep5	-0.22	spec_entropy1	-0.24
spec_entropy	-0.20	mcep2	-0.21	mcep18	-0.21	mcep7	-0.21
mcep6	-0.20	Hamm_breathy	-0.21	str4	-0.21	mcep6	-0.20
mcep11	-0.20	spec_entropy5	-0.20	mcep21	-0.19	mcep2	-0.19
...	...	...	...	...	...	...	...
mcep0	0.17	str4	0.21	voicing_rate	0.18	Hamm_effort	0.20
formant_disp	0.18	logpow	0.22	mcep1	0.20	pitch_entropy	0.21
pitch_entropy	0.19	Hamm_effort	0.23	B4	0.22	str1	0.22
voicing_rate	0.21	str3	0.24	spec_entropy4	0.27	mcep0	0.27

The higher positive and negative loadings of the PCA analysis are presented in Table 2. For PC1 mostly spectral features are the more loaded and also voicing rate. For PC2 spectral features, voicing strengths and voice quality features are highly loaded. Is interesting to notice that prosody features did not appear as good discriminators of the two conditions. An analysis of variance of these measures (one way ANOVA) indicates that almost all the measures are significantly different between polite and efficient condition with p-value < 0.001 except for str4, mcep6 and Hamm_effort on the male data.

4.2 SFFS-LM analysis

In Table 3 the features that best predict task accuracy for male and female data are presented. In this case all the data was used irrespective of the difference

score. If task accuracy is systematically affected by the politeness/efficiency condition we would like to know whether there are predominant acoustic features in each condition. In this case task accuracy in the polite condition seem to be better predicted by prosody features like max_f0, min_f0, std_f0, energy, and also some spectral features. Task accuracy in the efficiency condition seems to be less dependent on prosody features. An analysis of variance of these measures showed that most of these measures are not significantly different between the two classes polite and efficient. Here again the spectral features are more significantly different among the two conditions.

Table 3: Main acoustic predictors of accuracy for all the data. In parentheses is indicated the prediction error for each case. p-value after ANOVA of measures between the two classes polite and efficient is indicated by the significance codes: ***<0.001, **<0.01, *<0.05, . <0.1, ∘ <1.

Predicted accuracy Female				Predicted accuracy Male			
Polite (14.3%)		Efficient (13.95%)		Polite (4.85%)		Efficient (5.15%)	
mcep23	***	str2	∘	std_f0	**	mcep13	∘
max_f0	.	spec_entropy1	***	min_f0	*	std_f0	∘
spec_entropy2	**	spec_entropy2	**	energy	***	energy	***
min_f0	∘	mag2	∘	mcep10	∘	mcep4	***
mag1	*	mcep23	***	mcep18	***	spec_entropy4	***
std_f0	∘	min_f0	∘	mcep0	∘	str	∘
pitch_entropy	.	pitch_entropy	.	mcep6	***	str5	***
spec_entropy1	***	str1	***	pitch_entropy	∘	min_f0	*
mcep1	∘	max_f0	∘	str	∘	mcep7	***
str1	***	mcep5	***	mcep16	***	mcep10	∘

5 Conclusions

In this paper we have presented an acoustic analysis of politeness and efficiency in a cooperative time-sensitive task experiment.

In the PCA experiment we have found not so clear clusters or tendencies on the data analysed, although some individual sessions present clear clusters. One explanation could be that actually for some speakers there is no acoustic difference between the two conditions. In that case it would be necessary to perceptually annotate the words in the sessions so we can be sure that at perception level some words sound more polite than others in a more neutral condition.

In the SFFS-LM experiment we have found that task accuracy in the polite condition is better predicted by prosody features and task accuracy in the efficient condition seems to be less dependent on prosody features. This result seems to be more in line with the general tendency described on the literature that pitch is a good predictor of politeness [1, 5]. However, the analysis of variance of the features that better predict task accuracy showed that these features

do not discriminate well among the two conditions polite and efficient. So we can not conclude that the politeness condition was the only (or main) factor that affected task accuracy. One hypothesis, that will be analysed in future work, is that in the experiment task accuracy would have been also affected by task or cognitive load.

During the maze task, the trials in a block became increasingly more difficult, and the second and fourth blocks have the cursor controls flipped. The participants were informed about this change so they have to concentrate more on these blocks. In the literature it has been reported that speech rate, energy contour, F0 and spectral parameters are correlated with task load and stress [6], so we will analyse whether these features discriminate different levels of task load among the four blocks of the experiment.

Acknowledgements. The research leading to these results has received funding from the EU Programme FP7/2007-2013, under grant agreement no. 231287 (SSPNet).

References

1. Brown, P., Levinson, S.C.: Politeness some universals in language usage. Cambridge University Press (1987)
2. Brunet, P., Charfuelan, M., Cowie, R., Schröder, M., Donnan, H., Douglas-Cowie, E.: Detecting politeness and efficiency in a cooperative social interaction. In: Proc. Interspeech. Makuhari, Japan (2010)
3. Charfuelan, M., Schröder, M.: The vocal effort of dominance in scenario meetings. In: Proc. Interspeech. Florence, Italy (2011)
4. Douglas-Cowie, E., Cowie, R., Sneddon, I., Cox, C., Lowry, O., McRorie, M., Martin, J., Devillers, L., Abrilian, S., Batliner, A., Amir, N., Karpouzis, K.: The HUMAINE database: Addressing the collection and annotation of naturalistic and induced emotional data. In: Affective Computing and Intelligent Interaction, pp. 488–500 (2007)
5. Grawunder, S., Winter, B.: Acoustic correlates of politeness: prosodic and voice quality measures in polite and informal speech of korean and german speakers. In: Proc. Speech Prosody 2010. Chicago, Illinois, USA (2010)
6. Scherer, K.R., Grandjean, D., Johnstone, T., Klasmeyer, G., Bänziger, T.: Acoustic correlates of task load and stress. In: Proc. Interspeech. ISCA, Denver, Colorado, USA (2002)
7. Schöder, M., McKeown, G.: Considering social and emotional artificial intelligence. In: Proc. AISB 2010 Symposium "Towards a Comprehensive Intelligence Test". Leicester, UK. (2010)
8. Soman, V., Madan, A.: Social signaling: Predicting the outcome of job interviews from vocal tone and prosody. In: Proc. IEEE ICASSP. Dallas, Texas, USA (2010)
9. Vinciarelli, A., Salamin, H., Pantic, M.: Social signal processing: Understanding social interactions through nonverbal behavior analysis. In: IEEE Computer Vision and Pattern Recognition Workshops. pp. 42–49 (2009)
10. Wang, N., Johnson, W.L., Mayer, R.E., Rizzo, P., Shaw, E., Collins, H.: The politeness effect: Pedagogical agents and learning gains. Frontiers in Artificial Intelligence and Applications 125, 686–693 (2005)