Relationships Between Prosodic-Linguistic Features and High-Level Descriptors of Speed Dates

Sidd Jagadish, Ranjay Krishna, Gabriele Carotti-Sha

[1]Department of Statistics, Stanford University
[2]Department of Computer Science, Stanford University
[3] Symbolic Systems, Stanford University

siddj@stanford.edu, rak248@stanford.edu, gcarotti@stanford.edu

Abstract

We extract lexical and prosodic features of speech from 1493 speed dates at Stanford University. Each speed date is accompanied by information about what each participant thought of the other, including 1-10 assessments of courteousness and funniness. Here, we investigate prediction of these assessments from lexical and prosodic features. We find that for many labels, lexical and prosodic features can powerfully predict our high-level descriptors.

Index Terms: Prosody, Speed Date, AdaBoost, Gender differences in prosody, factor analysis, courteous, funny

1. Introduction

This template can be found on the conference website. Please use either a LibreOffice, MS-Word® or a LaTeX format file when preparing your submission. Information for full paper submission is available on the conference web site.

2. Data

We were given, courtesy of Jurafsky (2013), a dataset consisting of 1980 (CHECK THIS VALUE) heterosexual speed dates. For each date, we have two .wav files corresponding to the microphones attached to each participant. We also have high-level descriptors of each participant, including their height, weight, and ethnicity. In addition to this, for each participant, we have 1-10 assessments of various qualities of both themselves and each other person they went on a speeddate with, including funniness and courteousness.

2.1. Feature Extraction

Prosodic features were extracted using openSMILE. Lexical features were extracted in Python, with the help of the LIWC dictionary.

3. Exploratory Analysis

Before attempting to classify speakers as funny or not, we examine the underlying dimensions along which the data varies. To do, so we use exploratory factor analysis. To determine the number of factors, we use a scree test and various non-graphical measures, including parallel analysis and an optimal coordinates test, as described in (CITE). We find an optimal number of factors $k = 5$ for both males and females, conducted separately. Figure 1 shows the two scree plots.

Interestingly, although we find that the various factors for male and female speech are similar, they explain different amounts of variation in the data. For males, the first factor reflects high intensity values and low intensity variation, explaining

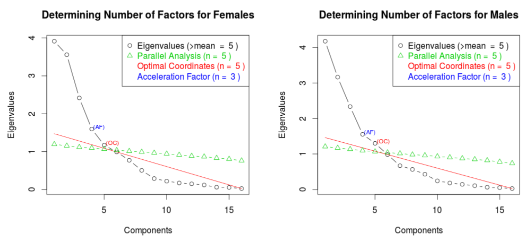

Figure 1: *Determining Number of Factors*

4. Page layout and style

Authors should observe the following rules for page layout. A highly recommended way to meet these requirements is to use a given template (LibreOffice, Word® or LaTeX) and check details against the corresponding example file.

4.1. Basic layout features

- Proceedings will be printed in A4 format. Authors must submit their papers in A4 format.
- Two columns are used except for the title part and possibly for large figures that need a full page width.
- Left margin is 20 mm.
- Column width is 80 mm.
- Spacing between columns is 10 mm.
- Top margin 25 mm (except for the first page which is 30 mm to the title top).
- Text height (without headers and footers) is maximum 235 mm.
- Headers and footers must be left empty (they will be added for printing and the INTERSPEECH 2013 media).
- Check indentations and spacings by comparing to this example file (in pdf).

4.1.1. Headings

Section headings are centered in boldface with the first word capitalized and the rest of the heading in lower case. Subheadings appear like major headings, except they start at the left margin in the column. Sub-sub-headings appear like sub-headings, except they are in italics and not boldface. See the examples given in this file. No more than 3 levels of headings should be used.

4.2. Text font

Times or Times Roman font is used for the main text. Font size in the main text must be 9 points, and in the References section 8 points. Other font types may be used if needed for special purposes. It is VERY IMPORTANT that while making the final PDF file, you embed all used fonts!

LaTeX users: users should use Adobe Type 1 fonts such as Times or Times Roman. These are used automatically by the interspeech2013.sty style file. Authors must not use Type 3 (bitmap) fonts.

4.3. Figures

All figures must be centered on the column (or page, if the figure spans both columns). Figure captions should follow each figure and have the format given in Figure **??**.

Figures should preferably be line drawings. If they contain gray levels or colors, they should be checked to print well on a high-quality non-color laser printer.

Graphics (i.e. illustrations, figures) must not use stipple fill patterns because they will not reproduce properly in Acrobat PDF. Please use only SOLID FILL COLORS.

Figures which span 2 columns (i.e. occupy full page width) should be placed at the top or bottom of the page.

4.4. Tables

An example of a table is shown as Table **??**. Somewhat different styles are allowed according to the type and purpose of the table. The caption text may be above or below the table.

Table 1: *This is an example of a table.*

ratio	decibels
1/1	0
2/1	≈ 6
3.16	10
10/1	20
1/10	-20
100/1	40
1000/1	60

4.5. Equations

Equations should be placed on separate lines and numbered. Examples of equations are given below. Particularly,

$$x(t) = s(f_\omega(t)) \tag{1}$$

where $f_\omega(t)$ is a special warping function

$$f_\omega(t) = \frac{1}{2\pi j} \oint_C \frac{\nu^{-1k} d\nu}{(1 - \beta\nu^{-1})(\nu^{-1} - \beta)} \tag{2}$$

A residue theorem states that

$$\oint_C F(z) dz = 2\pi j \sum_k Res[F(z), p_k] \tag{3}$$

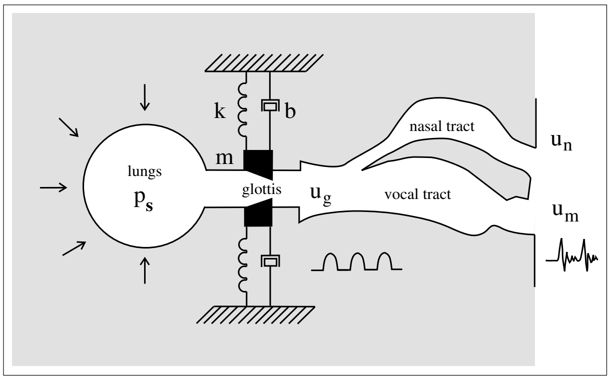

Figure 2: *Schematic diagram of speech production.*

Applying (**??**) to (**??**), it is straightforward to see that

$$1 + 1 = \pi \tag{4}$$

Finally we have proven the secret theorem of all speech sciences. No more math is needed to show how useful the result is!

4.6. Hyperlinks

For technical reasons, the proceedings editor will strip all active links from the papers during processing. Hyperlinks can be included in your paper, if written in full, eg. “http://www.foo.com/index.html”. The link text must be all black. Please make sure that they present no problems in printing to paper.

4.7. Multimedia files

The INTERSPEECH 2013 organizing committee offers the possibility to submit multimedia files. These files are meant for audio-visual illustrations that cannot be conveyed in text, tables and graphs. Just like you would when including graphics, make sure that you have sufficient author rights to the multimedia materials that you submit for publication. The proceeding media will NOT contain readers or players, so be sure to use widely accepted file formats, such as MPEG, Windows WAVE PCM (.wav) or Windows Media Video (.wmv) using standard codecs. The files you submit will be accessible from the abstract cards on the media and via a bookmark in the manuscript. From within the manuscript, refer to a multimedia illustration by its filename. Use short file names without blanks.

4.8. Page numbering

Final page numbers will be added later to the document electronically. *Don’t make any footers or headers!*

4.9. References

The reference format is the standard IEEE one. References should be numbered in order of appearance, for example [**?**], [**?**], and [**?**].

4.10. Abstract

The total length of the abstract is limited to 200 words. The abstract included in your paper and the one you enter during web-based submission must be identical. Avoid non-ASCII characters or symbols as they may not display correctly in the abstract book.

4.11. Author affiliation

Please list country names as part of the affiliation for each country.

4.12. Submitted files

Authors are requested to submit PDF files of their manuscripts. You can use commercially available tools or for instance http://www.pdfforge.org/products/pdfcreator. The PDF file should comply with the following requirements: (a) there must be no PASSWORD protection on the PDF file at all; (b) all fonts must be embedded; and (c) the file must be text searchable (do CTRL-F and try to find a common word such as 'the'). The proceedings editors (Causal Productions) will contact authors of non-complying files to obtain a replacement. In order not to endanger the preparation of the proceedings, papers for which a replacement is not provided timely will be withdrawn.

5. Discussion

This is the discussion. This is the discussion. This is the discussion. Is there any discussion.

This is the next paragraph of the discussion. And the last sentence of it.

6. Conclusions

Authors must proof read their PDF file prior to submission to ensure it is correct. Authors should not rely on proofreading the Word file. Please proofread the PDF file before it is submitted.

7. Acknowledgements

The ISCA Board would like to thank the organizing committees of past INTERSPEECH conferences for their help and for kindly providing the template files.

8. References

[1] Smith, J. O. and Abel, J. S., "Bark and ERB Bilinear Transforms", IEEE Trans. Speech and Audio Proc., 7(6):697–708, 1999.

[2] Soquet, A., Saerens, M. and Jospa, P.,"Acoustic-articulatory inversion", in T. Kohonen [Ed], Artificial Neural Networks, 371-376, Elsevier, 1991.

[3] Stone, H.S., "On the uniqueness of the convolution theorem for the Fourier transform", NEC Labs. Amer. Princeton, NJ. Online: http://citeseer.ist.psu.edu/176038.html, accessed on 19 Mar 2008.