Prosodic Styles and Personality Styles: Are the Two Interrelated?

Brigitte Zellner Keller

Dept. Clinical Psychology & Rehabilitative Psychiatry,
University of Bern, Switzerland
IMM, Lettres, University of Lausanne, Switzerland
Brigitte.ZellnerKeller@imm.unil.ch

Abstract

The "individuation" of oral language - what makes a speaker different from another - is still largely an unknown territory [1], especially with respect to the individual and creative use of speech prosody. This pilot study raises fundamental, methodological and empirical issues concerning the relationship between speakers' prosodic styles and their personality profiles. Our preliminary results support the hypothesis of a relationship between prosodic styles and "personality style" as perceived by listeners.

1. Introduction

A speaker's prosody can be conceived as a sophisticated combination between a linguistic code (stressed and unstressed syllables, typical durations of certain sounds, etc.), a rapid and momentary expression of emotion (sadness, happiness), a transient internal state related to the communicative situation (pride, irony, command, etc.), and finally, the expression of a more stable state that is the usual psychological behaviour of the speaker (figure1).

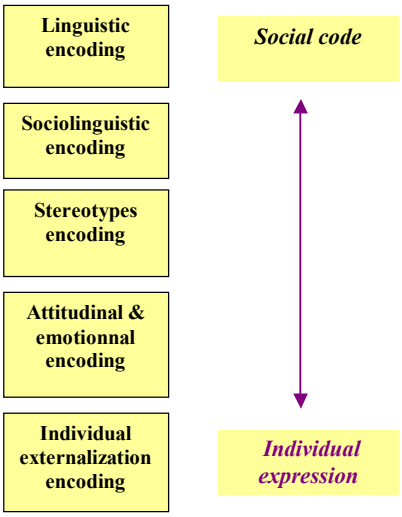

Figure 1. *Prosody, a multi-level encoding system*

Understanding the manner in which concepts and emotions are conveyed by means of oral expression, and what renders one speaker different from one another, has become a pressing issue in a variety of studies on verbal communication and spontaneous interactions [2], [3], [4], in studies on non-verbal communication - the non-linguistic component of speech prosody and body gestures [5], [6], [7], [8], as well as in psychology (e.g., [9], [10]). Paradoxically, these studies are limited from two points of view. First, they tend to focus on general tendencies, neglecting individual differences. Secondly, these studies tend to treat speech as a unique static and closed system, although all human beings need to be considered as open systems, subject to "meaningful modification" in environmental interaction [11]. For example, different prosodic styles are clearly associated with different communicative situations (dialogue, news report, lecture, etc.). Our study aims to better understand the stable component of prosody, that is, the "speaker's prosodic individuation" (see figure 2). In this sense, the aim of this pilot study is to explore the hypothesis of a relationship between prosody and psychological ratings.

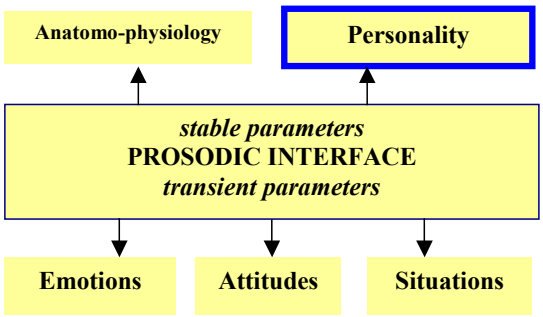

Figure 2. *The "speaker's prosodic individuation"*

2. Methodology

2.1. Recordings

Twelve native speakers of French (two women, ten men) were recorded in excellent laboratory conditions. The selected speakers were students or academic researchers of the University of Lausanne, with no communicative impairments. During a 15-min. recording session in which speakers had to perform various speech tasks, they were asked to tell a story from a cartoon presented on a sheet of paper. The cartoon is composed of five drawings, and every speaker had the same material from which to create the story. Depending on the speaker, stories contained between 51 to 145 words. A perceptual experiment as well as acoustic measurements were performed on the basis these speech samples. The recordings

were professionally digitized at 44.1 kHz, 16 bits. For segmentation and prosodic analysis, the recordings were resampled at 16 kHz after appropriate anti-alias filtering.

2.2. Perceptual Experiment

The perceptual experiment consisted in rating each speech sample according to an adaptation of a socio-psychological test, the "Impact Message Inventory" (IMI) [12], which is designed to measure a target person's interpersonal style. Each listener had to evaluate each speaker according to 15 interpersonal styles, on a scale ranging from 0 (little) to 7 (much). The twelve recordings were presented in random order to two groups of listeners, all students or academic researchers in psychology without communication impairments. The first group was composed of 28 French native speakers (6 men, 22 women), with a mean age of 24.9 and a mean comprehension of French of 6.96 on a self-report scale ranging from 0 (little) to 7 (native). The second group was composed of 12 German native speakers (9 men, 5 women), with a mean age of 37.7 and a mean comprehension of French of 3.41 on the same self-report scale. A database was built of all scores reported per speaker and listener.

2.3. Acoustic analysis

Acoustic analysis of the speech samples was conducted semi-automatically in Lausanne, thanks to computational tools developed at the LAIP Laboratory. Signals were automatically aligned with the transcribed text for segmentation and labelling (N=3898 phonetic segments), and labels were manually adjusted to insure temporal precision (interjudgemental agreement typically within 2-3 pitch cycles). These labels were used for the prosodic analysis. Automatic extraction furthermore furnished the following prosodic parameters: number of pauses and phones, f0 extraction on voiced segments, and intensity on vowels. A manual analysis provided measures of respiration, pauses, and speech rate. In addition, the proportion of f0 values (converted to semi-tones) and the proportion of intensity values out of a the mean speaker's registers were calculated. The final prosodic database contained a total of 21 parameters per speaker relating to various portions of the spoken utterances.

3. Results

Statistical analyses were performed with SPSS, v.11.0.

3.1. Comparison of the two groups of listeners

According to the Levene's test, the error variance of the dependent variable, the perceptual IMI scores, was equal across groups. A test for 2 x 15 cells (Group [Berne, Lausanne] by IMI items [15 items]) was applied. The ANOVA showed no significant differences between the two groups, despite the difference in levels of French comprehension: $F(2 , 330) = 1, 013$; $p = 0.315$, n.s.

3.2. IMI perceptual data

A factorial analysis of the 15 score categories x 12 speakers x 40 listeners suggested a three-component model. The Kaiser-Meyer-Olkin measure of sampling adequacy indicated a proportion of 0.815 of variance which might be accounted for by underlying factors. A factor model composed of three components resulting from this analysis explained 74.7 % of the variance (see figure 3). Component 1 grouped the following items: *dominant*, *selfish*, *hostile*, *intrusive*; component 2 grouped *mistrusting*, *detached*, *inhibited*, *submissive*, and component 3: *dependent*, *agreeable*, *supportive*, *sociable* and *cordial*. Three items were excluded from the model, since they were not significant at $p<0.05$: *docile*, *obliging*, and *intrusive*. Separate factorial analyses with each group of listeners produced similar models.

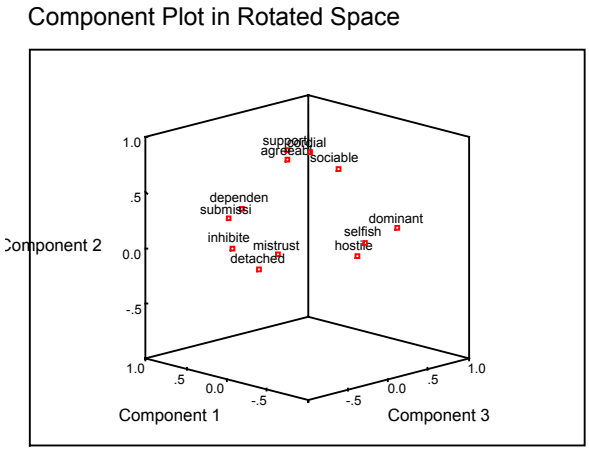

Figure 3. *Factorial model for the IMI perceptual ratings*

Coefficients of the factor analysis were saved and on this basis, a clustering of the speakers was applied, according to the Ward method.

3.3. Prosodic data

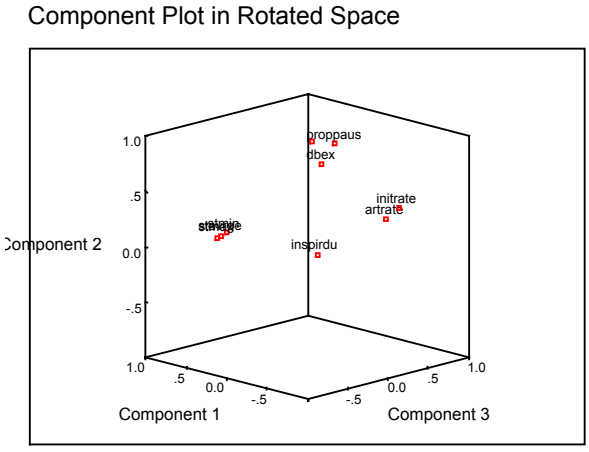

Figure 4**.** *Factorial model of the prosodic parameters*

A number of recent studies have suggested that men and women do not always use prosodic parameters in the same manner, in particular regarding the timing of phones, and also with respect to f0 and intensity parameters [see for example 13]. It was thus decided to exclude the data from the two women in our study from the analysis. F0 values were converted to semi-tones, a scale that is close to the perception

of pitch [14]. The Bartlett's test of sphericity indicates that significant relationships among prosodic variables were likely ($p<0.0001$). A factorial analysis for the 21 prosodic parameters of the 10 men suggested a 4-component model explaining 94.4 % of variance (figure 4). Component 1 contains the *f0 information*; component 2 contains the *silent pause information* plus the *intensity information*; component 3 contains the *speech rate information*; component 4 contains the *breathing information*.

Coefficients of the factor analysis were saved and on this basis, a clustering of the speakers was applied, according to the Ward method.

3.4. Mapping between the perceptual and the prosodic analysis

A significant correlation (Kendall's tau_b: 0.477, $p<0.05$) between "prosodic clustering" and "perceptual evaluation clustering" was found, suggesting that the prosodic indices furnished information directly relevant to the perceived personality style. Nine groups for ten speakers were identified by the correlational analysis.

Since nine groups for ten speakers does not represent a very strong clustering of prosodic and perceptual IMI data, a series of further cluster analyses was conducted and cluster memberships were saved. Table 1 (see end of article) recapitulates the results. The positive and negative signs show valences in the two types of clusters, the negative sign being attributed for low values and the positive sign for high values. Factor *f0* provides information about the height of the f0 register and its range. Factor *pauses* provides information on the length of silent pauses (with no breathing), the proportion of pause durations as compared to speech durations, and the control of intensity, expressed as the proportion of intensity values out of the mean speaker's intensity interval. Factor *rate* provides information on the speech rate at the beginning of the story and the mean articulation rate. Finally, factor *inspir* provides information on the mean duration of inspirations.

4. Discussion

We found a significant correlation between "prosodic clustering" and "perceptual evaluation clustering", suggesting that the prosodic indices furnished information directly relevant to the perceived personality style.To our knowledge, no study has systematically explored the relationship between prosodic parameters and perceived personality style. Yet it is a common observation that speakers, apart from their voices, differ with regard to their prosodic style. With respect to intonation and rhythm alone, one may describe a speaker's expression as fluent, lively, constrained, relaxed, etc. Intonation, rhythm and breathing are apparently the main parameters that convey these styles, and they contribute in some specific combination that still needs to be established. These are the parameters that are often imitated by impersonators, the imitation of the vocal component being more difficult to perform [15], [16]. Also, studies on twins are interesting in this context because of the speakers' morphological similarity [17]. For example, Loakes [18] found acoustic differences in the speech of twin pairs, despite clear perceptual similarity. In our study, four prosodic factors were identified: factor *f0*, factor *pauses* and *intensity*, factor *speech rate* and factor *breathing*.

4.1. Perception stability

In this pilot study, we raise the issue of a systematic relationship between the perception of speakers' oral communication and their personality styles. In the perceptual experiment, speech was neither degraded nor filtered, since a number of speech synthesis experiments have shown that perception changes considerably with degree of (un)naturalness of the stimulus. There is thus an intentional lack of control over the information taken into account by our listeners. It is possible that they made their judgements on the basis of a variety of features, among others, prosody. By testing French and German speaking listeners, we eliminated the hypothesis that lexical and syntactical material was of major relevance. (No significant differences were found between the two groups.) Prosody and voice quality are thus the most likely vehicles for speakers' personality styles.

Yet, another important issue to be clarified is the relationship between the "perceived personality style" and the personality type, as assessed by a direct personality measurement instrument.

4.2. Prosodic styles

The factorial model applied to the 21 prosodic parameters revealed that a speaker's prosodic style encapsulates a combination of information related to the speaker's f0 register, his use of silent pauses, combined with the use of abnormal intensity excursions, the initial speech rate and the mean articulation rate, as well as the mean duration of inspirations.

It appears in our data that speakers who tended to produce longer inspirations where those who used the silent pauses less actively and those who showed abnormal intensity excursions. Speakers who tended to produce shorter inspirations were those who had a faster initial speech rate as well as a faster mean articulation rate.

4.3. Relation between personality style and prosodic style

Despite the insufficient number of speakers for such a study, Table 1 shows an interesting consistency between psychological ratings and prosodic parameters. For example, three speakers evaluated as (- trusting, - social, + dominant) share 3 out of 4 of the prosodic parameters: few silent (non breathing) pauses and few intensity excursions, slow initial speech rate and slow mean articulation rate; long inspirations. The present results are in agreement with those found by Duez [19], where dominant speakers tend to show a good "control" of the temporal organisation of their spoken information with a slow speech rate and a higher use of breathing pauses. Further studies with larger groups of speakers should provide interesting further information for other personality styles.

5. Conclusion

In conclusion, considering the small number of subjects and the still relatively obscure relationship between speech production and speech perception, results are considered to be encouraging. This pilot study supports the hypothesis of a relationship between prosody and psychological ratings, and it suggests an extended study on these issues.

6. Acknowledgments

My deep gratitude goes to Wolfgang Tschacher (Berne), Klaus Scherer (Geneva), Veronique Aubergé (Grenoble), and Eric Keller (Lausanne) for their support and suggestions.

7. References

[1] Kienast, M., Glitza, F. (2003) Respiratory Sounds as an Idiosyncratic Feature in Speaker Recognition. Proceedings of 15th ICPhS, (1607-1610. Barcelona. ISBN 1-876346-48-5 © 2003 UAB.

[2] Grosz, B. & Sidner, B. (1986) Attention, Intentions, and the Structure of Discourse. Computational Linguistics, 12. 3, 175-204.

[3] Heeman, P.A., & Allen, J. F. (1999). Speech Repairs, Intonational Phrases and Discourse Markers: Modeling Speakers' Utterances in Spoken Dialog. Computational Linguistics 25(4), 527-571.

[4] Simon, A. C., & Auchlin, A. (2001). Multimodal, multifocal? Les hors-phase de la prosodie, in Cavé C., Guaïtella I. & Santi S. (éd.), Oralité et gestualité. Interactions et comportements multimodaux dans la communication, Actes du colloque ORAGE 2001, (629-633). Aix-en-Provence, 18-22 juin 2001, Paris, L'Harmattan.

[5] Douglas-Cowie, E., Cowie, R., Schröder, M. (2003). The Description of Naturally Occurring Emotional Speech. Proceedings of 15th ICPhS (2877-2881). Barcelona. ISBN 1-876346-48-5 © 2003 UAB.

[6] Aubergé V. & Cathiard M. (2002) The prosody of smile, Speech Communication Review, Special Speech and Emotion.

[7] Mc Neill, D. (2000). Language and Gesture. Cambridge University Press.

[8] Guaïtella I., Santi S., Cavé C., Bertrand R., Boyer J., Faraco M., Lagrue B., Mignard P., Paboudjian C., Purson A., 1998, Les relations voco-gestuelles dans la communication interpersonnelle, in : Santi S., Guaïtella I., Cavé C., Konopczynski G. (eds), ORAGE'98, ORAlité et Gestualité: communication multimodale, interaction, (13-25). Paris, L'Harmattan.

[9] Scherer, R. and Ekman, P. (1984). Approaches to emotion. Hillsdale, N.J. : L. Erlbaum Associates.

[10] Scherer, K. R. (1994). Interpersonal expectations, social influence, and emotion transfer. In P.D. Blanck (Ed.). Interpersonal expectations: Theory, research, and application (316-336). Cambridge and New York: Cambridge University Press.

[11] Feuerstein R. & Rynders, R.J., (1988). Helping Retarded People to Excel. London: Polonium, Press, 1988.

[12] Kiesler, D. J., Anchin, J. C., Perkins, M. J., Chirico, B. M., Kyle, E. M., & Federman, E. J. (1976). The Impact Message Inventory: Form II. Richmond: Virginia Commonwealth University.

[13] Simpson, A. P., Ericsdotter, C. (2003). Sex-Specific Durational Differences in English and Swedish. Proceedings of 15th ICPhS, (1113-1116). Barcelona. ISBN 1-876346-48-5 © 2003 UAB.

[14] Nolan, F. (2003). Intonational equivalence: an experimental evaluation of pitch scales. Proceedings of the 15th International Congress of Phonetic Sciences, 771-774, Barcelona.

[15] Besacier, L. (1998). Un modèle Parallèle pour la Reconnaissance Automatique du Locuteur. Thèse de doctorat.

[16] Zetterholm, E. (2002). A comparative survey of phonetic features of two impersonators. Fonetik, Volume 44, 129-132.

[17] Nolan, F., & Oh, T. (1996) Identical twins, different voices. Forensic Linguistics 3, 39-49

[18] Loakes, D. (2003) A Forensic Phonetic Investigation into the Speech Patterns of Identical and Non-Identical Twins. Proceedings of 15th ICPhS (691- 694). Barcelona. ISBN 1-876346-48-5 © 2003 UAB.

[19] Duez, D. (1991). La pause dans la parole de l''homme politique. Editions du CNRS,ISBN 2-222-04537-1.

Table 1. *Correspondence between IMI perception and prosodic parameters*

Speaker	Mistrusting	Social	Dominant	Prosody_F0	Prosody_pauses	Prosody_Rate	Prosody_inspir
2	-	-	+	-	-	-	+
3	+	+	-	-	+	-	+
4	-	-	+	+	-	-	+
5	-	+	-	+	+	-	-
6	+	-	-	-	+	+	-
7	-	-	+	-	-	-	+
8	-	+	-	+	-	+	-
9	+	-	-	+	-	-	+
10	+	-	+	-	-	-	-
12	-	+	+	+	-	-	+