

# THE IMPACT OF RHYTHMIC DISTORTIONS IN SPEECH ON PERSONALITY ASSESSMENT

*JAN VOLÍN*

Metropolitan University Prague  
jan.volin@ff.cuni.cz

*KRISTÝNA POESOVÁ*

Charles University in Prague  
kristyna.poesova@pedf.cuni.cz

*RADEK SKARNITZL*

Charles University in Prague  
radek.skarnitzl@ff.cuni.cz

## **Abstract**

The perennial question as to how perceived otherness in speech projects into listener assessment of one's personality has been systematically investigated within the field of foreign accentedness, vocal communication of affective states and vocal stereotyping. In the present study, we aimed at exploring non-native listeners' capacity to respond to differences in natural and modified native speech, particularly whether the manipulation of temporal structure in both stressed and unstressed syllables gives rise to any changes in the perception of the speaker's personality. The respondents' intuitive judgements were captured in the domain of the 'nervousness category' taken from the five-factor model of personality. Our results indicate an effect of temporal modifications on the listeners' judgements. Analysis of variance for repeated measures confirmed a highly significant shift of personality evaluations towards the undesired traits (e.g., nervousness, anxiety, querulousness). Several interesting interactions with the semantic contents of the utterances and with the intrinsic qualities of the speakers' voices were also found. We argue that the effects of accented speech go beyond conscious willingness to accept "otherness" and suggest a method for studying them.

**Keywords:** temporal structure, rhythm, neuroticism, perception, stereotype, foreign accent

## **1. Introduction**

With the global spread of English in today's vastly interconnected world and a growing number of non-native speakers using English in international contexts, we are witnessing a significant shift from the previous rejection of foreign-accentedness as something undesirable to the present greater openness to non-native speech. Calling for a wider acceptance of accent variation in the new millennium has replaced the hunt for accent

reduction common three decades ago. The firm adherence to native norms, British and American standards in particular, has been continually challenged along with other fundamental concepts operating within the traditional second language acquisition framework (Walker, 2010). These changes have resulted, apart from other things, in acknowledging both speakers and listeners as equally responsible for the success of communication instead of laying the main share of blame on the speaker, usually a non-native user, when a communication breakdown occurred. The idea of equal partnership in mutual interactions is believed to be implemented through raising general awareness about otherness in speech, or, more tangibly, through the development of supportive listening and phonological accommodation skills, especially on the part of native listeners. (Jenkins, Cogo & Dewey, 2011).

The demand that the listener take a more active role in the communicative process cannot be considered other than praiseworthy. However, without further investigation and deeper understanding of what impact foreign-accented speech has on communication in various settings and what kind of emotions it may stir up in the minds of interlocutors, its fuller acceptance may remain wishful thinking. The research carried out in this field so far has repeatedly confirmed the activation of dormant prejudices or creation of biases on the listener's part triggered by lower comprehensibility of foreign-accented speech and its increased processing load (Volín et al., 2013). This experience may consequently lead to downgrading non-native speakers' credibility, avoiding future interactions with them and other discriminatory acts (Munro & Derwing, 1995; Major, 2007; Lev-Ari & Keysar, 2010). Another strand of research approaches the discussed issue from the speaker's perspective and offers valuable insights into what it is like to speak differently and how the expected stigmatisation projects into the speaker's self-perception, sense of belonging (or anxiety due to experienced exclusion) and willingness to initiate or join conversations with native interlocutors (Gluszek & Dovidio, 2010; Gluszek et al., 2011).

It is truly amazing how much information a voice can reveal about its owner, although not all listener inferences correspond to true attributes of the speaker (Teshigawara, 2003). For example, it was found out that increased vocal attractiveness projects favourably into person perception. In other words, listeners think that people with attractive voices have more likeable and assertive personalities (Zuckerman & Driver, 1989). The vocal attractiveness stereotype offers a strong contrast to the negative reactions caused by the presence of foreign elements in one's speech and leaves researchers with a complicated task to shed some light on the nature of vocal stereotyping in different contexts.

Foreign-accentedness provided the first source of inspiration for the current study which scrutinizes the perception of otherness in speech, the second source being the relationship between prosodic variability and personality assessment. In his comprehensive overview, Scherer (2003) presents and discusses the major findings of research on vocal communication of emotion as well as the employed research paradigms. According to his classification, our experimental design ranks among inference studies which allow the determination of the relative effect of individual acoustic parameters on listener judgement either by cue masking or cue manipulation via synthesis (Scherer 2003, 239). A large number of works investigate the impact of global prosodic features on perceived affective states. The findings indicate, for example, that

increased speech rate signals excitement, agitation or anger while reduced speech rate is associated with qualities such as calm, content, irked (Kehrein, 2002), or that narrow and wide pitch range give rise to sadness and annoyance respectively (Scherer, 2003). Pitch level and loudness belong to other typically explored variables (Trouvain et al., 2006). Our focus turned to more local features, specifically to temporal distortions of vocalic nuclei in stressed and unstressed syllables, which seem to be examined rather sporadically to the best of our knowledge.

From Scherer's complex typology of affective states we chose personality traits for our investigation which the author characterizes as emotionally laden, stable personality dispositions and behaviour tendencies (Scherer 2003, 243). For personality assessment we applied the well-established and largely used five-factor model (see, e.g., Mohammadi et al., 2010; Trouvain et al., 2006) the consistency and adequacy of which with respect to everyday language were empirically verified by McCrae & Costa (1987). In order to narrow down the list of measured personality features and to gain better control of the design of the experiment, the factor of neuroticism was arbitrarily selected and the relevant adjectives – anxious, nervous, shy, open, sensitive and impulsive – created the basis for devising the perception instrument. We formulated the research question as follows: How does the local temporal distortion in speech mirror in listener perception of the speaker's nervousness? We hypothesized that temporal modifications would induce a change in personality evaluation.

## 2. Method

Our data were amassed via a perception test piloted by Volínová (2013), which was revised and expanded for the purposes of this study. The spoken material comprised 24 sentences devoid of emotionally charged language. The audio clips were from various radio programmes, presentations or read books, and lasted approximately 5 seconds each. They contained speech of 14 different English professional speakers. The local temporal manipulations were carried out in the Praat programme (Boersma & Weenink, 2012), using the PSOLA algorithm. The manipulations involved halving and doubling the duration of vowels in stressed and unstressed positions, respectively. This approach was inspired by the findings of Volín (2005) and Volín & Poesová (2008) in Czech-accented English. The aim was to create minor, yet perceptible alterations in the rhythmical structure of the utterances without any disturbing artefacts. Both original and resynthesized versions were randomized in the perception instrument to avoid the order effect. The potential threat of memory effect was reduced by desensitization in the form of short stretches of music separating individual test items and by the inclusion of fillers, specifically the voices of four BBC newsreaders, which were not manipulated and not included in the subsequent analyses.

The test was administered individually in a sound-treated studio of the Institute of Phonetics in Prague over the period of three weeks. The recordings were played via an Acer laptop and a set of headphones. The total number of listeners involved in the perception testing was 45. The majority of them were undergraduate students of the English language at the Faculty of Education, Charles University in Prague. The rest attended the Faculty of Arts of the same university. Their age ranged from 20 to 30

years, and none of the 39 females and 6 males reported any hearing disorder. All the respondents spoke English at the B2-level of CEFR or higher (upper intermediate use of English). They expressed their own voluntary interest to take part in the testing and received no financial compensation.

Reactions to the spoken stimuli were obtained by means of a battery of statements stemming from McCrae & Costa's pair adjectives subsumed within the neuroticism factor (1987). These statements either directly contained one of the key adjectives (e.g., *This person is anxious. This person is calm.*), or they described the typical behaviour of a person with the target attribute (e.g., *Watching a sad movie makes this person cry easily, This person easily strikes up a conversation with a stranger while waiting at a bus stop*). The behavioural type of statements prevailed as they reflect the real life inference processes more adequately. The identical statement was always matched with the natural and modified recordings of the same speaker. The whole test was divided into 4 blocks each comprising 14 statements, out of which 2 functioned as distractors. Statements were balanced so that positive (desired) and negative (avoided) human attributes were not biased towards the left or the right side of the answer sheet. Thus, although the seven-point scale was always oriented in the same way, the desired and undesired human features switched sides. Naturally, before the analyses, all the scores were adjusted so that undesired qualities were scored by negative numbers, while the desired (e.g., calmness, self-confidence) by positive values.

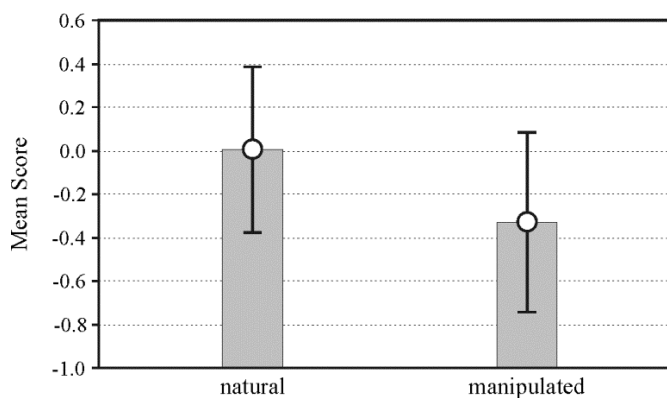
Precise instructions were crucial for the smooth running of the test. The respondents were first reassured that their knowledge of English was not the focus of testing. Subsequently, they were asked to try to capture the first impression of the speakers regardless of the content of the utterances. The testees were further informed that they would be given a few seconds to read each statement in advance and that the corresponding recording would be played only once in order to spark off their immediate reactions to the spoken material. Their perceptions were marked on a seven-point scale placed under each statement containing three positive values, zero and three negative values. The respondents had to choose one value that best reflected the extent to which they thought the statement about the speaker was true. They were asked to tick the zero value as a last resort and not to pay any attention to the meaning of the uttered sequences. The perception test lasted approximately 20 minutes including short breaks between the individual blocks.

### 3. Results

Each of the 48 target items was assessed by 45 listeners. This produced 2,160 evaluations or judgements. The evaluations were not uniform – on the one hand, the distribution of the scores over the 7-point scale was not random (rectangular distribution), on the other hand, individual judgements were sometimes quite spread and conflicted with each other. The standard deviations from the mean score given to an item ranged between 1.16 and 2.07 score points. The mean standard deviation for the natural items (i.e., non-manipulated) was 1.83 points, while the same measure for the manipulated items was 1.77 points. This difference was found insignificant and can be

interpreted by stating that the level of coherence with which our listeners judged the personality of the speaker was roughly equal for both the natural and manipulated items.

Figure 1 shows the difference between the grand mean scores awarded to natural and manipulated items. It is obvious that across all the speakers and items the mean score for the non-manipulated speech is about zero (more exactly 0.03), which, indeed, makes them on average neutral. The manipulated items yielded the mean score of  $-0.35$ , indicating more neuroticism for rhythmically distorted speech. One-way ANOVA for repeated (matched) measures was used to find out whether the difference was generalizable. The outcome suggests a high significance of the result:  $F(1, 23) = 14.2$ ;  $p < 0.001$ .



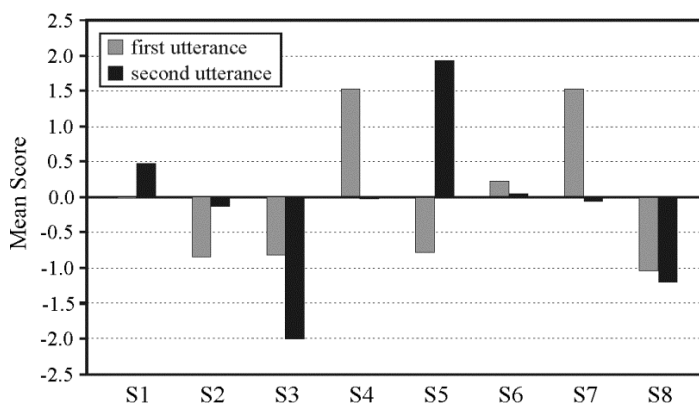
**Figure 1:** The grand mean scores given to natural (on the left) and rhythmically manipulated (on the right) items.

Table 1 displays the results for individual pairs of items. It can be observed that the shift in perception towards greater neuroticism occurred in 19 out of the 24 pairs of items, whereas in the remaining five pairs the shift happened in the opposite direction. Also, the items that contributed to the resulting overall trend (see above) produced larger individual differences between the natural and modified member of the pair ( $\Delta$  sc.) than the five deviant pairs.

Pair	$\Delta$ sc.	Pair	$\Delta$ sc.	Pair	$\Delta$ sc.	Pair	$\Delta$ sc.
1	-1.000	7	-0.667	13	-0.333	19	-0.156
2	-0.867	8	-0.600	14	-0.289	20	0.022
3	-0.867	9	-0.600	15	-0.267	21	0.156
4	-0.800	10	-0.422	16	-0.267	22	0.244
5	-0.689	11	-0.378	17	-0.222	23	0.600
6	-0.689	12	-0.356	18	-0.178	24	0.622

**Table 1:** Differences ( $\Delta$  sc.) between the mean scores of pairs of matched items. Ordered by magnitude. Negative values mean shift toward neuroticism.

Since eight of our fourteen voices were used more than once in the listening test, we were able to compare the ‘twin pairs’. Figure 2 shows the mean score which was attributed to the neutral (non-manipulated) items when the speaker occurred the first and second time in the test. The results indicate that the listeners’ evaluation of neuroticism was not based on purely intrinsic properties of the speakers’ voices. Only speakers S6 and S8 received very similar scores for both of their utterances. However, different aspects of neuroticism were queried for each of their utterances (e.g., anxiety and impulsiveness for S6). Contrary to that, S1 and S7 were queried for the same aspect, but the scores they received for their two utterances clearly differ from each other.



**Figure 2:** Mean score for eight speakers (S1 to S8) who occurred twice in the test with different utterances.

#### 4. Discussion

The research carried out in the field of foreign accent repeatedly confirms the impact of foreign features on the listeners and ultimately on the speakers themselves. Up to date, however, it has been relatively common to depict them holistically, as an aggregate. It is quite rare to find an experiment with personality perception scaling which assesses the accent attributes analytically from the phonetic point of view or isolates and weighs them against each other. Our study makes a modest step in this direction.

The manipulations or durational ratios of stressed and unstressed vowels affected the listeners’ ideas about the personality of the speaking individual. Although we took care not to make the durational changes too conspicuous, let alone distracting, they were still perceived and caused an increase of the neuroticism factor in the image of the speaker. The minds of the listeners used the information about the temporal structure of the utterance to portray the speaking individual.

We were also careful not to sway the respondents’ feeling by the content of the utterance and we explicitly asked them to ignore the meaning of the words spoken, but nevertheless the same voice uttering different words did not produce the same results. Our analyses do not permit an unambiguous interpretation of this finding. The meaning of the words could be an obvious answer, but it is also possible that subliminal local

features of the speaker's performance swayed the assessment. After all, one voice does make different impressions on different occasions. It could also have been the number or position of stressed syllables in one or the other utterance, which we did not control for. Similarly, the differences in the size of the resulting effect could have been influenced by an absolute change brought by the manipulation or by subtleties in the questionnaire wordings. All of these variables have to be investigated in future research, since the transparency of the methodology can stimulate comparative research in more than one accent of more than one language. This, in turn, would provide clearer answers concerning the effects of foreignness in speech and lead to more responsible decisions about social issues related to it.

### Acknowledgement

The authors would like to thank the 45 volunteers who underwent perceptual testing and provided their evaluations of the speakers' voices.

### References

- Boersma, P. and D. Weenink. 2012. Praat: doing phonetics by computer (version 5.3.14). Retrieved from <http://www.praat.org/>.
- Gluszek, A. and J. F. Dovidio. 2010. Speaking With a Non-native Accent: Perceptions of Bias, Communication Difficulties, and Belonging in the United States. *Journal of Language and Social Psychology* 29: 224–234.
- Gluszek, A., Newheiser, A. K. and J. F. Dovidio. 2011. Social Psychological Orientations and Accent Strength. *Journal of Language and Social Psychology* 30: 28–45.
- Jenkins, J., Cogo, A. and M. Dewey. 2011. State-of-the-Art Article: Review of developments in research into English as a lingua franca. *Language Teaching* 44: 281–315.
- Kehrein, R. 2002. The prosody of authentic emotions. *Proceedings of Speech Prosody: Aix-en-Provence, France*.
- Lev-Ari, S. and B. Keysar. 2010. Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology* 46: 1093–1096.
- Major, R. C. 2007. Identifying a foreign accent in an unfamiliar language. *Studies in Second Language Acquisition* 29: 539–556.
- McCrae, R. R. and P. T. Costa. 1987. Validation of the Five-Factor Model of Personality across the instruments and observers. *Journal of Personality and Social Psychology* 52: 81–90.
- Mohammadi, G., Vinciarelli, A. and M. Mortillaro. 2010. The Voice of Personality: Mapping Nonverbal Vocal Behavior into Trait Attributions. *Social Signal Processing Workshop: Firenze, Italy*.
- Munro, M. and T. Derwing. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45: 73–97

- Scherer, K. R. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication* 40: 227–256.
- Teshigawara, M. 2003. *Voices in Japanese Animation: A Phonetic Study of Vocal Stereotypes of Heroes and Villains in Japanese Culture*. Unpublished dissertation thesis.
- Trouvain, J., Schmidt, S., Schröder, M., Schmitz, M. and W. J. Barry. 2006. Modelling personality features by changing prosody in synthetic speech. *Proceedings of Speech prosody: Dresden, Germany*.
- Volín, J. 2005. Rhythmical properties of polysyllabic words in British and Czech English. In: J. Čermák et al. (Eds.) *Patterns, A Festschrift for Libuše Dušková*. Praha: *Kruh moderních filologů*: 279–292.
- Volín, J. and Poesová, K. 2008. Temporal and spectral reduction of vowels in English weak syllables. In: A. Grmelová et al. (Eds.), *Plurality and Diversity in English Studies*. Praha, UK PedF: 18–27.
- Volín, J., Weingartová, L. and R. Skarnitzl. 2013. Spectral Characteristics of Schwa in Czech Accented English. *Research in Language* 11 v1: 31–39. DOI: 10.2478/v10015-012-0008-6
- Volínová, S. 2013. *Perceptual Stereotypes in Social Interaction*. Unpublished bachelor thesis.
- Walker, R. 2010. *Teaching the Pronunciation of English as a Lingua Franca*. Oxford: Oxford University Press.
- Zuckerman, M. and R. E. Driver. 1989. What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior* 13: 67–82.