Do Sound Segments Contribute to Sounding Charismatic? Evidence from a Case Study of Steve Jobs’ and Mark Zuckerberg’s Vowel Spaces

Oliver Niebuhr
University of Southern Denmark, Mads Clausen Institute, Alson 2, DK-6400 Sønderborg, Denmark.

Simon Gonzalez
Centre of Excellence for the Dynamics of Language, The Australian National University, 13 Ellery Crescent, Canberra ACT 2601.

(Received 24 June 2018; accepted 22 November 2018)

The paper presents a case study of two popular US American CEOs. It compares the acoustic vowel space sizes of the more charismatic speaker Steve Jobs and those of the less charismatic speaker Mark Zuckerberg, as part of an initial acoustic step to examine a traditional claim of rhetoric that clearer speech makes a speaker sound more charismatic. Analysing about 2,000 long and short vowel tokens from representative keynote speech excerpts of the two speakers shows that Jobs’ vowel space is, across various segmental and prosodic context factors, significantly larger than that of Zuckerberg, whose vowel space is strongly reduced particularly when addressing investors. The differences in vowel-space size are consistent with the claim of rhetoric that a clear articulation is a key characteristic of a charismatic speaker. The discussion of the results describes further experimental steps required to back up the link between clear pronunciation and speaker charisma.

1. INTRODUCTION

1.1. The Phonetics of Charismatic Speech

Spoken language is not just the exchange of propositions. On the contrary, it is in the first place a social action, and “this fact both shapes the nature of the activity and its consequences.”26 We use speech for expressing our emotions and sharing them with others, as well as for influencing the thoughts and actions of others. To this extent, charisma as “the art of persuasion”65 through “emotion-laden leader signalling”51 is indeed a core element of spoken language — and its phonetic essence is surprisingly little understood.

It is against this background that Rosenberg and Hirschberg called for an empirical definition of charisma in speech.53 They analysed the acoustic-prosodic characteristics of male US politicians and related them to perceived charisma. Their analysis led to the conclusion that higher levels of fundamental frequency (F0), intensity, and speaking rate, as well as a larger F0 range, make speakers sound more charismatic. These findings were consistent with analyses of other political leaders in the US and in Europe.9,25,61,65 Moreover, the same prosodic strategy also works for business leaders,42,43 except that females are more likely to lower rather than raise their F0 level.46 Furthermore, Niebuhr et al. added to the picture that shorter prosodic phrases, larger numbers of emphatic pitch accents, high-energy voices (higher values of %V, spectral emphasis, and HNR1), and more variable speech rhythm (higher VarcoV values) also support a speaker’s charismatic impact.42,43,45

While an empirical definition of charismatic prosody is within reach (at least for Western Germanic languages and/or the Western culture), a whole area of speech has hardly been addressed so far: sound segments.5 Manuals on rhetoric and leadership have claimed ever since that clear and crisp articulation of “every phrase and word”30 “is imperative to develop charisma.”7

Basically, this claim makes sense from the perspective of basic ethological principles like the Effort Code.21 According to the Effort Code, a fundamental behavioural pattern of all biological organisms, they spend more time and effort on things and actions that are more important to them. In order to understand the implications of this basic principle in human everyday life, one only needs to think about how elaborate the table is set when important guests are coming as compared to how simple the table setting is when one eats alone. If a simple table setting were used for important guests, the implicit message would be that the host does not care about his/her actions and/or that the invited guests are not important to the host. The same is true for speech communication. Investing more effort into articulating sound segments would indicate from the Effort Code’s point of view that the conveyed message is important and that the speaker shows appreciation for his/her audience. In contrast, mumbling would implicitly signal that the speaker does not care about his/her message and the audience as well.

Similarly, the Hypo-Hyper (H&H) theory of Lindblom regarded a clear, effortful articulation (hyper-speech) as being listener-oriented, with, for example, an aim to meet the relatively higher frequencies. Voices with a perceived high “volume” and “power” lose less energy towards higher frequencies. HNR stands for harmonics-to-noise ratio and quantifies the ratio (in dB) between the periodic energy and the noise energy of a signal at a given point in time.

---

1 %V is the average proportion of vowel segments in an utterance. It was introduced as a rhythm measure, but since vowels are the most energetic parts of an utterance, %V is also highly correlated with the perceived “volume” and “power” of a voice.50 Spectral emphasis refers to the difference between the total acoustic energy of the signal at a given point in time and the energy in the lower frequency region of that signal (0-1.5*median, following Traummüller and Eriksson).56 Thus, spectral emphasis quantifies the loss of energy towards higher frequencies. Voices with a perceived high “volume” and “power” lose less energy towards higher frequencies. HNR stands for harmonics-to-noise ratio and quantifies the ratio (in dB) between the periodic energy and the noise energy of a signal at a given point in time.