MELODY IN SPEECH: THE ROLE OF HARMONY PRINCIPLES IN ASSESSING PERSONALITY TRAITS

Dalida Valeeva

Chair of Speech Communication, Institute of Language & Communication, TU Berlin, Hardenbergstr. 16-18, 10623 Berlin, Germany

Abstract

This research investigates the connection between voice characteristics and personality traits, proposing that shared principles of harmony underlie both music and speech perception. Personality was assessed using both self-report and external ratings based on the NEO Five-Factor Inventory. The perceptual evaluation involved analysing three types of speech samples: a sustained vowel sound [a], a reading passage, and spontaneous speech. To deepen auditory insights, acoustic features of the recordings were examined using the software PRAAT, identifying specific vocal markers tied to individual personality dimensions. The speech signals were also transformed into musical notation via Neuratron AudioScore Professional to analyse pitch intervals and distinguish between harmonic and disharmonic transitions. Findings revealed strong correlations - particularly involving extraversion and neuroticism - between vocal parameters and personality. Key indicators included pitch (F0), pitch variability, jitter, shimmer, harmonics-to-noise ratio (HNR), energy distribution, articulation clarity, and consonance/dissonance in speech melody. Notably, introverted or emotionally unstable individuals tended to show higher values in F0, jitter, shimmer, and HNR, whereas emotionally stable and extroverted individuals displayed greater pitch variability and more energy in higher frequencies. Features like articulation precision and melodic pitch transitions also played vital roles in distinguishing personality profiles. The study enhances our understanding of vocal expression's influence on perception and suggests applications in speech analysis technology, recruitment processes, psychological counselling, and career guidance.

Keywords: *vocal cues of personality, speech melody, prosodic features, personality traits in the voice, music in speech*

1. INTRODUCTION

The human voice carries a wealth of information about a speaker, including gender, background, age, social status, physical and emotional health, and personality traits. Despite rapid advancements in technology and the growing role of artificial intelligence, the human ear remains unparalleled in interpreting vocal cues. This study explores how personality is reflected through vocal expression, with a particular focus on the musical elements of speech perception. Although voice and speech have long been subjects of scientific inquiry, the intersection between music and spoken language - and how they are jointly processed - remains underexplored. Drawing on the overlap in signal generation and processing between speech and music, the study suggests that universal principles of harmony apply not only to musical perception but also to how we perceive voices and spoken communication.

1.1. Acoustic indicators of personality dimensions

The human voice is a central means by which individuals express their identity - communicating gender, origin, age, social status, health, emotions, and personality. Even with advances in AI and technology, the human ear remains unmatched in detecting subtle cues in vocal communication, often outperforming software in real-world scenarios.

This perceptual skill is evolutionarily rooted: both humans and animals use vocal signals to convey traits like body size, social standing, and intentions. Remarkably, this sensitivity develops before birth - human foetuses can recognize and respond to their mothers' voices even in utero [1].

Historically, voice has long been interpreted as a reflection of character. Ancient Greek rhetoricians believed vocal traits signified moral and emotional qualities - a tense, deep voice denoted courage, while a high, weak one signaled cowardice. Variations in pitch and timbre were thought to reveal irritability, mockery, vanity, and other dispositions.

Scientific interest in the connection between personality and voice gained momentum in the 1930s, especially in Germany and England. For instance, vocal assessments played a role in military officer selection prior to World War II, where qualities like warm tone and melodic phrasing were linked to empathy, and harsh, monotonous voices to willpower [2].

Pioneers like Paul J. Moses [3] explored how emotional instability affects vocal features, noting patterns such as shaky pitch, inconsistent tone, and insecure intonation among neurotic individuals. Fährmann [4] further analysed speech characteristics alongside personality profiles, finding that hesitant speakers exhibited uneven tempo, quiet delivery, and erratic stress - whereas dominant personalities demonstrated strong articulation and increased speech tempo.

A more structured and empirical approach was introduced by Klaus R. Scherer [5], who validated personality-related vocal markers across cultures. His research showed that listeners could accurately identify personality traits - like extraversion - purely from voice samples, with pitch and vocal intensity being key indicators. Stable and extroverted voices were typically resonant, warm, and loud, while deeper voices tended to be associated with emotional stability [6]. This framework inspired subsequent studies - such as Müller's [7] investigation into pitch (F0) and formants - which revealed that lowering F0 influenced perceptions of neuroticism and extraversion. Notably, deeper male voices were rated as more competent and agreeable, while formant adjustments significantly altered impressions of female voices, enhancing perceptions of altruism and helpfulness.

Later research further supported these findings, including studies by McAleer et al. [8], Sendlmeier [9], Belin et al. [10], and Pearsell et al. [11], each contributing nuanced insights into how acoustic features correlate with psychological traits.

1.2. Perception of music and speech

Speech and musicality are uniquely human abilities. While language's role in daily life is more apparent, music holds a fundamental place in all cultures across the globe. It often marks both the beginning and end of auditory experience - a mother's lullaby heard by a newborn, and the final sounds perceived by individuals in a coma [12].

Music and language share cognitive foundations, relying heavily on pitch perception. Both activities engage similar neural mechanisms: Broca's area, for example, processes structural elements of speech and identifies both verbal slips and musical inaccuracies. Brain activity measurements show that off-key notes in music and incorrect words in a sentence elicit comparable neurological responses [13].

Jourdain [14] notes that the brain can distinguish around 1,300 tones within the musical range, though these are grouped into a limited number of categories. The relative pitch between tones is more important than their exact frequency - a concept that also applies to phoneme recognition, where numerous variations of a sound can be understood without perfect articulation. This form of categorical perception is crucial for both spoken language and music.

A longstanding question in music cognition involves the distinction between consonant and dissonant intervals. According to Helmholtz [15], how consonant an interval sounds depends on the difference in frequencies between the tones. Simpler frequency ratios result in more consonant intervals. However, the boundary between consonance and dissonance is shaped by historical and cultural contexts. In Western music theory (17th - early 20th century), intervals like the octave, fifth, thirds, and sixths were classified as consonant, whereas seconds, sevenths, the tritone, and sometimes the fourth were deemed dissonant, depending on stylistic interpretation [16].

Research into the musical nature of speech is still emerging. Raven [17] examined whether emotional meaning in speech is conveyed using musical-like intervals in pitch contours. By analysing key

inflection points, the study found that dissonant intervals were commonly used to express negative emotions - such as fear, anger, sadness, boredom, and disgust - each linked to specific tonal changes.

Ploug & Niebuhr [18] explored how pitch intervals affect speaker charisma. Their findings revealed that less charismatic individuals tend to favor smaller intervals like minor and major seconds and minor thirds. Conversely, more charismatic speakers were found to use broader, more stable intervals - namely perfect fourths, fifths, and major sixths.

2. METHODOLOGY

2.1. Methods

The study aimed to investigate whether a relationship exists between vocal characteristics and personality assessments, and to identify which traits are expressed through speech and vocal delivery. To accomplish this, the alignment between self-rated personality dimensions and traits perceived auditorily by others was analysed.

The assessment involved 11 native German speakers who completed a self-evaluation, and 28 listeners who rated the speakers' personalities using the NEO-Five Factor Inventory framework [19]. The listening test incorporated three types of speech samples: a sustained vowel sound [a] lasting 6 seconds, a passage from Aesop's fable The North Wind and the Sun, and spontaneous speech describing a typical weekly routine - each lasting 15–17 seconds. Statistical analysis was conducted to evaluate the ratings. Mean values for perceived traits were calculated for each speaker, and correlations between self-evaluation and listener perceptions were examined.

To complement auditory analysis, acoustic features were measured using the PRAAT software [20], allowing identification of vocal markers linked to personality traits. Additionally, the speech recordings were transcribed into musical notation using Neuratron AudioScore Professional [21], enabling the study of pitch transitions and the distinction between harmonic and disharmonic intervals in speech patterns.

Fig. 1. A spoken utterance converted into musical notation using Neuratron AudioScore Professional [21]

3. RESULTS

Analysis of the perception test data revealed a notable correlation between self-assessed personality traits and those perceived auditorily by unrelated listeners. Specifically, neuroticism demonstrated a strong alignment between self-evaluation and external perception, while extraversion showed a moderate but statistically significant correlation. The other traits - openness to experience, conscientiousness, and agreeableness - did not yield significant results.

This pattern was consistent across the combined results of all three speech samples as well as in their individual analyses. Neuroticism correlated in both the read text and spontaneous speech, whereas extraversion showed alignment in the sustained vowel recording. No meaningful correlations emerged for the remaining personality traits when examined separately.

The subsequent acoustic analysis aimed to identify speech signal parameters that may underpin perceptions of neuroticism and extraversion. From the pool of 11 speakers, two were selected based on their contrasting personality scores: female speaker 03, who registered the highest neuroticism and lowest extraversion across all speech types, and female speaker 06, who represented the opposite end of the spectrum. Each speech sample type - vowel, read text, and spontaneous utterance - was analysed individually.

In the vowel subtest, key phonation metrics such as fundamental frequency (F0) and the first and second formants (F1 and F2) were evaluated. For the read and spontaneous speech samples, the focus shifted to prosodic and articulatory features.

F0 emerged as a critical marker in vocal expression of personality. Using PRAAT software, average F0 values (equivalent to perceived pitch), F0 range, and standard deviation (reflecting pitch variability) were measured. Speaker 03 exhibited the highest average F0 and a relatively flat pitch contour, resulting in a monotonous vocal impression. By contrast, Speaker 06 displayed greater pitch variation, lending her speech a more dynamic and expressive quality.

	Speaker **03**	Speaker **06**	Speech Type
F0 Range	141,5 Hz	200,8 Hz	read
	170,2 Hz	226,5 Hz	spontaneous
Standard deviation	28, 36 Hz	44,74 Hz	read
	29,96 Hz	52,61 Hz	spontaneous

Table 1. F0 variation-relevant values of the most unstable & introvert person (Speaker 03) and the most stable and extrovert one (Speaker 06)

In terms of frequency instability, notable differences were observed in prosodic markers such as jitter - small fluctuations in pitch (F0) - and shimmer, which reflects variations in amplitude between successive sound waves. These parameters varied significantly between emotionally stable and unstable speakers. For example, Speaker 03 exhibited jitter and shimmer values that were twice as high as those of Speaker 06, resulting in a clearly perceptible tremble in her voice. It's also important to highlight that elevated levels of jitter and shimmer are commonly associated with emotional states such as fear, boredom, and anger, which tend to manifest acoustically through increased vocal irregularities.

Vowel [a]	Speaker **03**	Speaker **06**
Jitter	1,9 %	0,5 %
Shimmer	4,4 %	2,6 %

Table 2. Jitter and shimmer values in sustained vowel [a] of the speakers with the most contrasting neuroticism and extraversion values

Further analysed phonation differences concerned the harmonics-to-noise ratio (HNR): the most unstable person had the highest coefficient. This result supports previous studies that consider this coefficient as a measurable parameter for marking arousal, fear, and fatigue. During the recording, care was taken to ensure that the volume remained relatively consistent for all speakers. The volume level on the DAT recorder was adjusted accordingly. The intensity values differ little among the speakers with opposite neuroticism and extraversion values. However, the energy distribution shows the following trend: stable and extroverted speakers have more energy in the higher frequencies than unstable and introverted ones. This result also confirms previous studies.
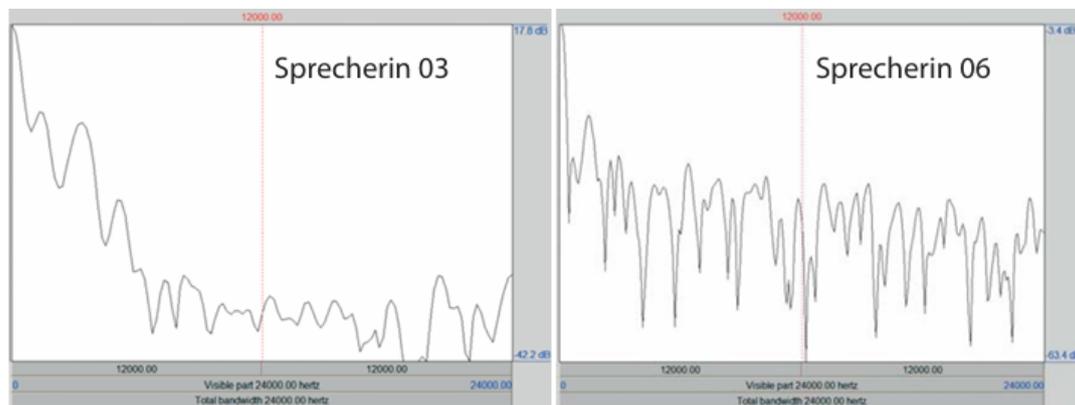

Fig. 2. Comparison of the spectral slices of Speaker 03 and Speaker 06 (spontaneous speech).

The emphasis patterns of the speakers also differ. For example, Speaker 06 generally uses a pitch accent, while Speaker 03 mainly emphasizes by lengthening the vowels. This, among other things, has the effect of slowing down the speaking tempo. Additionally, the pitch accent of Speaker 06 contributes to a greater variation in the pitch contour.

The analysis of the formant charts of the two speakers led to the clear conclusion that the vowels of the more stable person were more decentralized, resulting in a clearer articulation of the vowels. The more unstable and introverted speaker, on the other hand, articulated the vowels more centrally and with less mouth opening (indicated by the lower F1 values). This tendency has already been observed in the sustained vowel. The formant bands of Speaker 06 showed a clearer and more stable structure than those of Speaker 03. However, the accuracy of articulation was reversed for consonants: Speaker 03 articulates the consonants more clearly and precisely, with less assimilation and reduction compared to Speaker 06. For example, the sound combinations [ts] + [ʃ] or [n] + [t] were pronounced separately and with little reduction by Speaker 03, in contrast to Speaker 06. Additionally, Speaker 03 tended to devoice voiced consonants: [d] becomes [t], [b] sounds like [p], [z] like [s]. In contrast, Speaker 06 shows the opposite tendency: voiceless consonants are voiced, for example, [t] sounds like [d]. The plosives had a stronger burst, and a longer voice-onset time compared to Speaker 06. The results regarding articulation characteristics were similar in both read and spontaneous utterances.

Additionally, the music psychology-based analysis revealed that subconscious recognition and differentiation of consonant versus dissonant speech melodies play a meaningful role in evaluating

personality traits such as neuroticism and extraversion. In the speech of the individual rated as the most introverted and emotionally unstable, dissonant intervals - especially the minor second, one of the most dissonant intervals - were predominant. Her melodic speech contour displayed a consistently descending pattern with minimal variation.

By contrast, the person perceived as the most extroverted and emotionally stable produced speech containing more consonant intervals, notably minor thirds and fourths. Her melody contour was shaped like a bell curve, showing a wider pitch range and greater variation. These findings support the conclusion that principles of musical perception - particularly how we process melodic intervals - are highly relevant for understanding and analysing speech.

4. DISCUSSION

Speaker 03 was evaluated as both emotionally unstable and introverted, suggesting that the observed vocal traits and speech behaviours may contribute simultaneously to assessments of neuroticism and extraversion. To disentangle these dimensions more clearly, it would be necessary to include a participant who scores highly in one trait - such as introversion - but remains emotionally stable. A larger, more diverse group of speakers and listeners could also yield more representative and reliable findings.

Another important consideration is whether the verbal content of the spontaneous speech influences personality ratings. In this study, that influence appears minimal. Despite similar weekend descriptions (e.g., "meeting friends," "going out"), participants were judged quite differently on neuroticism and extraversion. This highlights the dominant role of vocal delivery and speaking style in shaping personality perception.

Interestingly, although Speaker 03 reported average levels of neuroticism and extraversion on the NEO-FFI self-assessment, she was perceived by listeners as unstable and introverted. She later disclosed that she was experiencing emotional distress due to relationship issues during the recording, which may have affected her vocal expression.

The acoustic analysis focused on female voices due to their distinct personality assessment divergence. To gain a fuller understanding of voice-based personality judgments, further research involving male speakers is essential. Additionally, the study's acoustic findings should be empirically validated through future testing.

A promising direction for follow-up research could involve AI-based manipulation of vocal parameters at both segmental and suprasegmental levels. For instance, pitch contours could be adjusted to explore effects of consonance and dissonance in speech melody, followed by listener-based perception tests to compare resulting personality assessments.

5. CONCLUSIONS

The study found that listeners' ratings closely matched speakers' self-assessments for neuroticism - and to a lesser but still significant extent for extraversion - while no reliable agreement emerged for openness, conscientiousness, or agreeableness. Acoustic measures showed that the most neurotic/introverted voice featured higher mean pitch (F0), jitter, shimmer, and HNR with less variation, whereas the most stable/extroverted voice had greater pitch variability, more energy in higher frequencies, and clearer vowel dispersion. Analysis of speech melodies further revealed that dissonant intervals (especially minor seconds) dominated in speech reflecting emotional instability, while consonant intervals (minor thirds and fourths) prevailed in verbal expression of extroverted personalities, confirming that musical harmony principles apply to speech perception. These results highlight how vocal and melodic cues shape personality impressions and point to the value of larger, gender-balanced samples and AI-driven prosody manipulations for future validation and application in voice-based assessment tools.

ACKNOWLEDGMENTS

REFERENCES

1. Carvalho, ES, Rosado de Miranda Justo, JM, Gratier, M & Ferreira Rodrigues da Silva, HM 2019, "The Impact of Maternal Voice on the Fetus: A Systematic Review", *Current Women's Health Reviews*, pp. 196–206.

2. Kreiman, J & Sidtis, D 2011, "Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception", Wiley-Blackwell, UK, p. 342.

3. Moses, P 1956, "Die Stimme der Neurose", Thieme Verlag, Stuttgart, p. 107.

4. Fährmann, R 1967, "Die Deutung des Sprechausdrucks: Studien zur Einführung in die Praxis der charakterologischen Stimm- und Sprechweise", Bouvier Verlag, Bonn, p. 74.

5. Scherer, KR 1979, "Social markers in speech", Cambridge University Press, Cambridge, p. 157.

6. Scherer, KR 1982, "Vokale Kommunikation: Nonverbale Aspekte des Sprachverhaltens", Beltz Verlag, Weinheim, p. 198.

7. Müller, R 2009, "Stimme und Persönlichkeit", Franzbecker Verlag, Hildesheim/Berlin, pp. 106–119.

8. McAleer, P, Todorov, A & Belin, P 2014, "How do you say 'Hello'? Personality impressions from brief novel voices", *PLOS ONE*, vol. 9, no. 3, e90779.

9. Sendlmeier, WF 2019, "Sprechwirkungsforschung: Grundlagen und Anwendungen mündlicher Kommunikation", Logos Verlag, Berlin, *Bd. 10*: *Mündliche Kommunikation*, pp. 82, 277.

10. Belin, P, Boehme, B & McAleer, P 2017, "The sound of trustworthiness: Acoustic-based modulation of perceived voice personality", *PLOS ONE*, vol. 12, no. 10, e0185651.

11. Pearsell, S & Pape, D 2023, "The effects of different voice qualities on the perceived personality of a speaker", *Frontiers in Communication*, vol. 7, 909427.

12. Cordes, I 2005, "Der Zusammenhang kultureller und biologischer Ausdrucksmuster in der Musik", Lit Verlag, Münster, p. 19.

13. Maess, B, Koelsch, S, Gunter, TC & Friederici, AD 2001, "Musical syntax is processed in Broca's area: an MEG study", *Nature Neuroscience*, vol. 4, no. 5, pp. 540–545.

14. Jourdain, R 1998, "Das wohltemperierte Gehirn", Spektrum Akademischer Verlag, Heidelberg/Berlin, pp. 94–101.

15. Helmholtz, HV 1863, "Die Lehre von den Tonempfindungen", F. Vieweg & Sohn, Braunschweig, pp. 372–373.

16. Ebeling, M 2008, "Konsonanz und Dissonanz", Bruhn, H, Kopiez, R & Lehmann, AC (eds), *Musikpsychologie*, Rowohlt Taschenbuch Verlag, Reinbek bei Hamburg, pp. 499–500.

17. Raven, H, Bartels, A & Sendlmeier, WF 2005, "Gemeinsame Kommunikationsstrategien von Sprache und Musik – Musikalische Intervalle im Grundfrequenzverlauf emotionaler Äußerungen", Sendlmeier, WF & Bartels, A (eds), *Stimmlicher Ausdruck in der Alltagskommunikation, Bd. 4: Mündliche Kommunikation*, Logos Verlag, Berlin, pp. 109–132.

18. Ploug, M & Niebuhr, O 2019, "There is music in speech melody! – How pitch intervals shape speaker charisma", *Proceedings of the 1st International Seminar on the Foundations of Speech: Pausing, Breathing and Voice*, Sonderborg, Denmark, pp. 68–76.

19. Borkenau, P & Ostendorf, F 1993, "NEO-Fünf-Faktoren-Inventar (NEO-FFI): Handanweisung", Hogrefe, Göttingen.

20. Boersma, P & Weenink, D 1992–2023, "Praat: doing phonetics by computer", Version 6.4.01, http://www.praat.org/.

21. Neuratron Ltd. 2023, "Neuratron AudioScore Professional", Version 8.9, https://download.cnet.com/audioscore-ultimate/.