Definitions of voice naturalness from the literature

Deviation-based naturalness

1. Abur et al. (2021):

"and naturalness was defined as conforming to the listener's standards of rate, rhythm, intonation, and stress patterning and to the syntactic structure of the utterance being produced." (page 4687)

2. Anand & Stepp (2015):

"Speech naturalness can be described as how the speech of a person with a speech disorder compares with that of typical speech or, in the case of an acquired disorder, how an individual's speech compares to its premorbid state" (page 1134)

3. Coughlin-Woods et al. (2005):

"Speech Naturalness is a term defined from the listener's perspective as speech that sounds normal or natural and allows the listener's attention to focus on the message rather than the speech pat tern" (page 295)

4. Eadie & Doyle (2002):

"Naturalness was defined as a perceptually derived, overall description of prosodic adequacy. That is, speech was defined as natural if it conforms to the listener's standards of rate, rhythm, intonation, and stress pattern, and if it conforms to the syntactic structure of the utterance being produced" (page 1091)

5. Eadie et al. (2008):

"Speech samples were defined as natural if [they] conform to the listener's standards of rate, rhythm, intonation, and stress pattern" (page 47)

6. Jones et al. (2019):

"How speech compares to that encountered in typical speech." (page 329)

7. Klopfenstein (2015):

"Speech is natural if it conforms to the listener's standard of rate, rhythm, intonation and stress patterning, and if it conforms to the syntactic structure of the utterance being produced'' (page 938)

8. Klopfenstein (2016):

"[…] sounding 'natural' may be speaking like she or he did before the onset of dysarthria, even if such a goal is out of reach." (page 124)

9. Klopfenstein et al. (2020):

"Speech is natural if it conforms to the listener's standard of rate, rhythm, intonation and stress patterning, and if it conforms to the syntactic structure of the utterance being produced" (page 327)

10. Lehner et al. (2022):

"It describes to what extent the way someone speaks sounds natural and not irritating, i.e., whether it conforms to the expected standard of unimpaired speech" (page 990)

11. Merritt & Bent (2020):

“Speech naturalness has been conceptualized as a listener’s scaling between an individual’s speech and the listener’s representation of typical speech patterns, including the dimensions of rate, rhythm, intonation, stress patterning, and syntactic structure” (page 2054)

12. Moya-Galé et al. (2024):

“[…] naturalness, which has been defined as the degree to which speech meets the typical patterns in terms of intonation, voice quality, rate, rhythm, and intensity, with respect to the syntactic structure of the utterance” (page 2)

13. Nussbaum et al. (2023):

“By naturalness, we understand the voice stimulus to be perceived as a plausible outcome of the human speech production system” (page 1)

14. Schoelderle et al. (2023):

“Speech naturalness refers to a rather broad perceptual impression representing the overall quality of a person’s speech output in relation to what is conceptualized as normal or natural” (page 1633/1634)

15. Venkatraman & Sivasanka (2018):

“How likely are you to hear a speaker like this, during a typical day” (page 1541)

16. Vogel et al. (2019):

“[…] degree to which individuals sound ‘different’ from healthy peers” (page 1265)

17. Yorkston et al. (1990):

“Natural speech was defined for this study as conventional in terms of intonation, voice quality, rate, rhythm, or intensity adjustments. Unnatural or bizarre speech, on the other hand, was defined as markedly deviating from the expected or conventional pattern.” (page 551)

Human-likeness-based naturalness

1. Baird et al. (2018):

"Human likeness has been used […] to describe how accurately the machine is able to imitate a human." (page 2864)

2. Gong & Nass (2007):

"Humanoid is adopted as the working term in this article to refer to the state of being humanlike but bearing the clear artificiality of computer synthesis." (page 164)

3. Hu & Lu (2021):

"Humanness perception of technology is defined as the degree to which a user feels a certain technology or system is human-like (versus machine-like)." (page 3)

4. Hyppa-Martin et al. (2024):

"Naturalness refers to whether synthetic speech is perceived as uniquely human, despite being computer-generated." (page 5)

5. Im et al. (2023):

"[…] perceived naturalness is a valid way to capture the overall quality of synthetic voice and that people associate perceived naturalness of voice with humanlikeness of voice." (page 3)

6. Li et al. (2023):

"Voice human-likeness, referring to the level of naturalness and likeness to human voice, is one of the fundamental anthropomorphic features of robots." (page 2)

7. Lu et al. (2021):

"Anthropomorphism, which is the assignment of a human form, characteristics, or behavior to nonhuman objects […]." (page 1)

8. Mawalim et al. (2022):

"natural speech is the speech most closely perceived as a human voice." (page 10)

9. Mayo et al. (2011):

"[…] that is, with regard to how much like real speech each utterance seemed to be." (page 316)

10. Rodero (2017):

"[…] naturalness is the quality by which a voice sounds more similar to human speech." (page 336)

11. Velner et al. (2020):

"[…] an interaction that has the same conversational flow as humans have interacting with each other." (page 570)

12. Yamasaki et al. (2017):

"Naturalness refers to how closely the output sounds like human speech." (page 389.e1)

Combination of both

1. Duville et al. (2022):

"[…] acoustic tendencies that define naturalness from human to synthesized voices." (page 1) & "In sum, the neural specialization of the human brain to process speech is underlined by acoustic properties that are perceived as voice naturalness (i.e., speech intrinsic property to be recognized as a social ecological sound)." (page 2)

2. Kapolowicz et al. (2022):

"[…] voices which sound like they could come from an actual human being (which should be rated as more natural) and voices that sound more fictitious, such as a cartoon character or a monster (which should be rated as less natural)." (page 429)

3. Malisz et al. (2019):

"[..] the necessary cues to make synthesis realistic, e.g., natural, easily intelligible, and specific to a given speaker." (page 2)

Comments on naturalness, that were not counted as definitions

1. Assmann et al. (2006):

"Listeners were informed that they would hear a range of computer-generated voices varying in naturalness, and that some voices might sound like children or cartoon characters." (page 2)

2. Martin et al. (1984):

"Naturalness will not be defined for you." (page 54)

3. Meltzner & Hillman (2005):

"[...] which of the two tokens in each pair ''sounded more like normal natural speech'' (NNS)." (page 770)

4. Nusbaum et al. (1995):

"There is no extant objective definition of naturalness that we are aware of--it is a voice quality that is purely subjective." (page 8)

5. Ratcliff et al. (2002):

"Speech naturalness has been defined as speech output that sounds normal or natural to the listener." (page 11)