List of included papers focusing on naturalness in voices (alphabetical order):

1. Abdulrahman, A., & Richards, D. (2022). Is Natural Necessary? Human Voice versus Synthetic Voice for Intelligent Virtual Agents. Multimodal Technologies and Interaction, 6(7), 51. https://doi.org/10.3390/mti6070051
2. Abur, D., Subaciute, A., Daliri, A., Lester-Smith, R. A., Lupiani, A. A., Cilento, D., Enos, N. M., Weerathunge, H. R., Tardif, M. C., & Stepp, C. E. (2021). Feedback and Feedforward Auditory-Motor Processes for Voice and Articulation in Parkinson's Disease. J Speech Lang Hear Res, 64(12), 4682–4694. https://doi.org/10.1044/2021_JSLHR-21-00153
3. Anand, S., & Stepp, C. E. (2015). Listener Perception of Monopitch, Naturalness, and Intelligibility for Speakers With Parkinson's Disease. Journal of Speech, Language, and Hearing Research, 58(4), 1134–1144. https://doi.org/10.1044/2015_JSLHR-S-14-0243
4. Assmann, P. F., Dembling, S., & Nearey, T. M. (2006). Effects of frequency shifts on perceived naturalness and gender information in speech. In INTERSPEECH. Symposium conducted at the meeting of Citeseer.
5. Aylett, M. P., Vinciarelli, A., & Wester, M. (2020). Speech Synthesis for the Generation of Artificial Personality. IEEE Transactions on Affective Computing, 11(2), 361–372. https://doi.org/10.1109/TAFFC.2017.2763134
6. Baird, A., Jørgensen, S. H., Parada-Cabaleiro, E., Cummings, N., Hantke, S., & Schüller, B. (2018). The Perception of Vocal Traits in Synthesized Voices: Age, Gender, and Human Likeness. Journal of the Audio Engineering Society, 66(4), 277–285. https://doi.org/10.17743/jaes.2018.0023
7. Baird, A., Jørgensen, S. H., Parada-Cabaleiro, E., Hantke, S., Cummins, N., & Schuller, B. (2017). Perception of Paralinguistic Traits in Synthesized Voices. In G. Fazekas, M. Barthet, & T. Stockman (Eds.), Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences (pp. 1–5). ACM. https://doi.org/10.1145/3123514.3123528
8. Baird, A., Parada-Cabaleiro, E., Hantke, S., Burkhardt, F., Cummings, N., & Schüller, B. (2018, September 2). The Perception and Analysis of the Likeability and Human Likeness of Synthesized Speech. In Interspeech 2018 (pp. 2863–2867). ISCA. https://doi.org/10.21437/Interspeech.2018-1093
9. Birkholz, P., & Drechsel, S. (2021). Effects of the piriform fossae, transvelar acoustic coupling, and laryngeal wall vibration on the naturalness of articulatory speech synthesis. Speech Communication, 132, 96–105. https://doi.org/10.1016/j.specom.2021.06.002
10. Birkholz, P., Martin, L., Xu, Y., Scherbaum, S., & Neuschaefer-Rube, C. (2017). Manipulation of the prosodic features of vocal tract length, nasality and articulatory precision using articulatory synthesis. Computer Speech & Language, 41, 116–127. https://doi.org/10.1016/j.csl.2016.06.004
11. Cabral, J. P., Cowan, B. R., Zibrek, K., & McDonnell, R. (2017). The Influence of Synthetic Voice on the Evaluation of a Virtual Character. In Interspeech 2017 (pp. 229–233). ISCA. https://doi.org/10.21437/Interspeech.2017-325
12. Coughlin-Woods, S., Lehman, M. E., & Cooke, P. A. (2005). Ratings of speech naturalness of children ages 8-16 years. Perceptual and Motor Skills, 100(2), 295–304. https://doi.org/10.2466/pms.100.2.295-304
13. Diel, A., & Lewis, M. (2024). Deviation from typical organic voices best explains a vocal uncanny valley. Computers in Human Behavior Reports, 14, 100430. https://doi.org/10.1016/j.chbr.2024.100430

14. Duville, M. M., Alonso-Valerdi, L. M., & Ibarra-Zarate, D. I. (2022). Neuronal and behavioral affective perceptions of human and naturalness-reduced emotional prosodies. Frontiers in Computational Neuroscience, 16, 1022787. https://doi.org/10.3389/fncom.2022.1022787
15. Duville, M. M., Alonso-Valerdi, L. M., & Ibarra-Zarate, D. I. (2024). Improved emotion differentiation under reduced acoustic variability of speech in autism. BMC Medicine, 22(1), 121. https://doi.org/10.1186/s12916-024-03341-y
16. Eadie, T. L., & Doyle, P. C. (2002). Direct Magnitude Estimation and Interval Scaling of Naturalness and Severity in Tracheoesophageal (TE) Speakers. J Speech Lang Hear Res, 45(6), 1088–1096. https://doi.org/10.1044/1092-4388(2002/087)
17. Eadie, T. L., Doyle, P. C., Hansen, K., & Beaudin, P. G. (2008). Influence of speaker gender on listener judgments of tracheoesophageal speech. Journal of Voice, 22(1), 43–57. https://doi.org/10.1016/j.jvoice.2006.08.008
18. Ehret, J., Bönsch, A., Aspöck, L., Röhr, C. T., Baumann, S., Grice, M., Fels, J., & Kuhlen, T. W. (2021). Do Prosody and Embodiment Influence the Perceived Naturalness of Conversational Agents' Speech? ACM Transactions on Applied Perception, 18(4), 1–15. https://doi.org/10.1145/3486580
19. Euler, H. A., Merkel, A., Hente, K., Neef, N., Wolff von Gudenberg, A., & Neumann, K. (2021). Speech restructuring group treatment for 6-to-9-year-old children who stutter: A therapeutic trial. Journal of Communication Disorders, 89, 106073. https://doi.org/10.1016/j.jcomdis.2020.106073
20. Eyssel, F., Kuchenbrandt, D., Bobinger, S., Ruiter, L. de, & Hegel, F. (2012). 'If you sound like me, you must be more human'. In H. Yanco, A. Steinfeld, V. Evers, & O. C. Jenkins (Eds.), HRI' 12: Proceedings of the seventh annual ACM/IEEE Conference on Human-Robot Interaction : March 5-8, 2012 Boston, Massachusetts, USA (pp. 125–126). Association for Computing Machinery. https://doi.org/10.1145/2157689.2157717
21. Ferstl, Y., Thomas, S., Guiard, C., Ennis, C., & McDonnell, R. (2021). Human or Robot? Investigating voice, appearance and gesture motion realism of conversational social agents. In Proceedings of the 21th ACM International Conference on Intelligent Virtual Agents (pp. 76–83). ACM. https://doi.org/10.1145/3472306.3478338
22. Gong, L., & Nass, C. (2007). When a Talking-Face Computer Agent is Half-Human and Half-Humanoid: Human Identity and Consistency Preference. Human Communication Research, 33(2), 163–193. https://doi.org/10.1111/j.1468-2958.2007.00295.x
23. Goy, H., Kathleen Pichora-Fuller, M., & van Lieshout, P. (2016). Effects of age on speech and voice quality ratings. The Journal of the Acoustical Society of America, 139(4), 1648. https://doi.org/10.1121/1.4945094
24. Hardy, T. L. D., Rieger, J. M., Wells, K., & Boliek, C. A. (2020). Acoustic Predictors of Gender Attribution, Masculinity-Femininity, and Vocal Naturalness Ratings Amongst Transgender and Cisgender Speakers. Journal of Voice, 34(2), 300.e11-300.e26. https://doi.org/10.1016/j.jvoice.2018.10.002
25. Higgins, D., Zibrek, K., Cabral, J., Egan, D., & McDonnell, R. (2022). Sympathy for the digital: Influence of synthetic voice on affinity, social presence and empathy for photorealistic virtual humans. Computers & Graphics, 104, 116–128. https://doi.org/10.1016/j.cag.2022.03.009
26. Hu, P., Lu, Y., & Gong, Y. (2021). Dual humanness and trust in conversational AI: A person-centered approach. Computers in Human Behavior, 119, 106727. https://doi.org/10.1016/j.chb.2021.106727
27. Hyppa-Martin, J., Lilley, J., Chen, M., Friese, J., Schmidt, C., & Bunnell, H. T. (2024). A large-scale comparison of two voice synthesis techniques on intelligibility, naturalness, preferences, and attitudes toward voices banked by individuals with amyotrophic lateral

sclerosis. Augmentative and Alternative Communication, 40(1), 31–45. https://doi.org/10.1080/07434618.2023.2262032
28. Ilves, M., & Surakka, V. (2013). Subjective responses to synthesised speech with lexical emotional content: the effect of the naturalness of the synthetic voice. Behaviour & Information Technology, 32(2), 117–131. https://doi.org/10.1080/0144929X.2012.702285
29. Ilves, M., Surakka, V., & Vanhala, T. (2011). The Effects of Emotionally Worded Synthesized Speech on the Ratings of Emotions and Voice Quality. In (pp. 588–598). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-24600-5_62
30. Im, H., Sung, B., Lee, G., & Xian Kok, K. Q. (2023). Let voice assistants sound like a machine: Voice and task type effects on perceived fluency, competence, and consumer attitude. Computers in Human Behavior, 145, 107791. https://doi.org/10.1016/j.chb.2023.107791
31. Jones, H. N., Crisp, K. D., Kuchibhatla, M., Mahler, L., Risoli, T., Jones, C. W., & Kishnani, P. (2019). Auditory-Perceptual Speech Features in Children With Down Syndrome. American Journal on Intellectual and Developmental Disabilities, 124(4), 324–338. https://doi.org/10.1352/1944-7558-124.4.324
32. Kapolowicz, M. R., Guest, D. R., Montazeri, V., Baese-Berk, M. M., & Assmann, P. F. (2022). Effects of Spectral Envelope and Fundamental Frequency Shifts on the Perception of Foreign-Accented Speech. Language and Speech, 65(2), 418–443. https://doi.org/10.1177/00238309211029679
33. Klopfenstein, M. (2015). Relationship between acoustic measures and speech naturalness ratings in Parkinson's disease: A within-speaker approach. Clinical Linguistics & Phonetics, 29(12), 938–954. https://doi.org/10.3109/02699206.2015.1081293
34. Klopfenstein, M. (2016). Speech naturalness ratings and perceptual correlates of highly natural and unnatural speech in hypokinetic dysarthria secondary to Parkinson’s disease. Journal of Interactional Research in Communication Disorders, 7(1), 123–146. https://doi.org/10.1558/jircd.v7i1.27932
35. Klopfenstein, M., Bernard, K., & Heyman, C. (2020). The study of speech naturalness in communication disorders: A systematic review of the literature. Clinical Linguistics & Phonetics, 34(4), 327–338. https://doi.org/10.1080/02699206.2019.1652692
36. Ko, S., Barnes, J., Dong, J., Park, C. H., Howard, A., & Jeon, M. (2023). The Effects of Robot Voices and Appearances on Users’ Emotion Recognition and Subjective Perception. International Journal of Humanoid Robotics, 20(01), Article 2350001. https://doi.org/10.1142/S0219843623500019
37. Kühne, K., Fischer, M. H., & Zhou, Y. (2020). The Human Takes It All: Humanlike Synthesized Voices Are Perceived as Less Eerie and More Likable. Evidence From a Subjective Ratings Study. Frontiers in Neurorobotics, 14, 593732. https://doi.org/10.3389/fnbot.2020.593732
38. Lee, E.-J. (2010). The more humanlike, the better? How speech type and users’ cognitive style affect social responses to computers. Computers in Human Behavior, 26(4), 665–672. https://doi.org/10.1016/j.chb.2010.01.003
39. Lehner, K., & Ziegler, W. (2022). Clinical measures of communication limitations in dysarthria assessed through crowdsourcing: Specificity, sensitivity, and retest-reliability. Clinical Linguistics & Phonetics, 36(11), 988–1009. https://doi.org/10.1080/02699206.2021.1979658
40. Li, M., Guo, F., Wang, X., Chen, J., & Ham, J. (2023). Effects of robot gaze and voice human-likeness on users’ subjective perception, visual attention, and cerebral activity in voice conversations. Computers in Human Behavior, 141, 107645. https://doi.org/10.1016/j.chb.2022.107645
41. Lu, L., Zhang, P., & Zhang, T. (2021). Leveraging “human-likeness” of robotic service at restaurants. International Journal of Hospitality Management, 94, 102823. https://doi.org/10.1016/j.ijhm.2020.102823

42. Mackey, L. S., Finn, P., & Ingham, R. J. (1997). Effect of speech dialect on speech naturalness ratings: A systematic replication of Martin, Haroldson, and Triden (1984). Journal of Speech, Language, and Hearing Research, 40(2), 349–360. https://doi.org/10.1044/jslhr.4002.349
43. Malisz, Z., Henter, G. E., Valentini-Botinhao, C., Watts, O., Beskow, J., & Gustafson, J. (2020). Modern speech synthesis for phonetic sciences: a discussion and an evaluation. Center for Open Science. https://doi.org/10.31234/osf.io/dxvhc
44. Martin, R. R., Haroldson, S. K., & Triden, K. A. (1984). Stuttering and speech naturalness. The Journal of Speech and Hearing Disorders, 49(1), 53–58. https://doi.org/10.1044/jshd.4901.53
45. Mawalim, C. O., Galajit, K., Karnjana, J., Kidani, S., & Unoki, M. (2022). Speaker anonymization by modifying fundamental frequency and x-vector singular value. Computer Speech & Language, 73, 101326. https://doi.org/10.1016/j.csl.2021.101326
46. Mayo, C., Clark, R. A. J., & King, S. (2011). Listeners' weighting of acoustic cues to synthetic speech naturalness: A multidimensional scaling analysis. Speech Communication, 53(3), 311–326. https://doi.org/10.1016/j.specom.2010.10.003
47. McGinn, C., & Torre, I. (2019, March 11–14). Can you Tell the Robot by the Voice? An Exploratory Study on the Role of Voice in the Perception of Robots. In 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 211–221). IEEE. https://doi.org/10.1109/HRI.2019.8673305
48. Meltzner, G. S., & Hillman, R. E. (2005). Impact of Aberrant Acoustic Properties on the Perception of Sound Quality in Electrolarynx Speech. J Speech Lang Hear Res, 48(4), 766–779. https://doi.org/10.1044/1092-4388(2005/053)
49. Merritt, B., & Bent, T. (2020). Perceptual Evaluation of Speech Naturalness in Speakers of Varying Gender Identities. J Speech Lang Hear Res, 63(7), 2054–2069. https://doi.org/10.1044/2020_JSLHR-19-00337
50. Mitchell, W. J., Szerszen, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & Macdorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. I-Perception, 2(1), 10–12. https://doi.org/10.1068/i0415
51. Moore, B. C. J., & Tan, C.-T. (2003). Perceived naturalness of spectrally distorted speech and music. The Journal of the Acoustical Society of America, 114(1), 408–419. https://doi.org/10.1121/1.1577552
52. Moya-Galé, G., Pagano, G., & Walsh, S. J. (2024). Perceptual consequences of online group speech treatment for individuals with Parkinson's disease: A pilot study case series. International Journal of Speech-Language Pathology, 1–16. https://doi.org/10.1080/17549507.2024.2330538
53. Nusbaum, H. C., Francis, A. L., & Henly, A. S. (1997). Measuring the naturalness of synthetic speech. International Journal of Speech Technology, 2(1), 7–19.
54. Nussbaum, C., Pöhlmann, M., Kreysa, H., & Schweinberger, S. R. (2023). Perceived naturalness of emotional voice morphs. Cognition & Emotion, 1–17. https://doi.org/10.1080/02699931.2023.2200920
55. Parmar, D., Olafsson, S., Utami, D., Murali, P., & Bickmore, T. (2022). Designing Empathic Virtual Agents: Manipulating Animation, Voice, Rendering, and Empathy to Create Persuasive Agents. Autonomous Agents and Multi-Agent Systems, 36(1). https://doi.org/10.1007/s10458-021-09539-1
56. Rao M V, A., Victory J, S., & Ghosh, P. K. (2018). Effect of source filter interaction on isolated vowel-consonant-vowel perception. The Journal of the Acoustical Society of America, 144(2), EL95. https://doi.org/10.1121/1.5049510
57. Ratcliff, A., Coughlin, S., & Lehman, M. (2002). Factors influencing ratings of speech naturalness in augmentative and alternative communication. Augmentative and Alternative Communication, 18(1), 11–19. https://doi.org/10.1080/aac.18.1.11.19

58. Rodero, E. (2017). Effectiveness, attention, and recall of human and artificial voices in an advertising story. Prosody influence and functions of voices. Computers in Human Behavior, 77, 336–346. https://doi.org/10.1016/j.chb.2017.08.044
59. Rodero, E., & Lucas, I. (2023). Synthetic versus human voices in audiobooks: The human emotional intimacy effect. New Media & Society, 25(7), 1746–1764. https://doi.org/10.1177/14614448211024142
60. Romportl, J. (2014). Speech Synthesis and Uncanny Valley. In A. Horák, P. Sojka, I. Kopeček, & K. Pala (Eds.), Lecture Notes in Computer Science: Vol. 8655, Text, speech and dialogue (pp. 595–602). Springer International Publishing. https://doi.org/10.1007/978-3-319-10816-2_72
61. Sarigul, B., & Urgen, B. A. (2023). Audio–Visual Predictive Processing in the Perception of Humans and Robots. International Journal of Social Robotics, 15(5), 855–865. https://doi.org/10.1007/s12369-023-00990-6
62. Schölderle, T., Haas, E., & Ziegler, W. (2023). Speech Naturalness in the Assessment of Childhood Dysarthria. American Journal of Speech-Language Pathology, 32(4), 1633–1643. https://doi.org/10.1044/2023_AJSLP-23-00023
63. Schreibelmayr, S., & Mara, M. (2022). Robot Voices in Daily Life: Vocal Human-Likeness and Application Context as Determinants of User Acceptance. Frontiers in Psychology, 13, 787499. https://doi.org/10.3389/fpsyg.2022.787499
64. Seaborn, K., Miyake, N. P., Pennefather, P., & Otake-Matsuura, M. (2021). Voice in Human–Agent Interaction. ACM Computing Surveys, 54(4), 1–43. https://doi.org/10.1145/3386867
65. Tamagawa, R., Watson, C. I., Kuo, I. H., MacDonald, B. A., & Broadbent, E. (2011). The Effects of Synthesized Voice Accents on User Perceptions of Robots. International Journal of Social Robotics, 3(3), 253–262. https://doi.org/10.1007/s12369-011-0100-4
66. Urakami, J., Sutthithatip, S., & Moore, B. A. (2020). The Effect of Naturalness of Voice and Empathic Responses on Enjoyment, Attitudes and Motivation for Interacting with a Voice User Interface. In M. Kurosu (Ed.), Lecture Notes in Computer Science. Human-Computer Interaction. Multimodal and Natural Interaction (Vol. 12182, pp. 244–259). Springer International Publishing. https://doi.org/10.1007/978-3-030-49062-1_17
67. van Prooije, T., Knuijt, S., Oostveen, J., Kapteijns, K., Vogel, A. P., & van de Warrenburg, B. (2023). Perceptual and Acoustic Analysis of Speech in Spinocerebellar ataxia Type 1. Cerebellum (London, England). Advance online publication. https://doi.org/10.1007/s12311-023-01513-9
68. Velner, E., Boersma, P. P., & Graaf, M. M. de (2020). Intonation in Robot Speech. In T. Belpaeme, J. Young, H. Gunes, & L. Riek (Eds.), Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (pp. 569–578). ACM. https://doi.org/10.1145/3319502.3374801
69. Venkatraman, A., & Sivasankar, M. P. (2018). Continuous Vocal Fry Simulated in Laboratory Subjects: A Preliminary Report on Voice Production and Listener Ratings. American Journal of Speech-Language Pathology, 27(4), 1539–1545. https://doi.org/10.1044/2018_AJSLP-17-0212
70. Vogel, A. P., Stoll, L. H., Oettinger, A., Rommel, N., Kraus, E.-M., Timmann, D., Scott, D., Atay, C., Storey, E., Schöls, L., & Synofzik, M. (2019). Speech treatment improves dysarthria in multisystemic ataxia: A rater-blinded, controlled pilot-study in ARSACS. Journal of Neurology, 266(5), 1260–1266. https://doi.org/10.1007/s00415-019-09258-4
71. Yamasaki, R., Montagnoli, A., Murano, E. Z., Gebrim, E., Hachiya, A., Lopes da Silva, J. V., Behlau, M., & Tsuji, D. (2017). Perturbation Measurements on the Degree of Naturalness of Synthesized Vowels. Journal of Voice, 31(3), 389.e1-389.e8. https://doi.org/10.1016/j.jvoice.2016.09.020

72. Yorkston, K. M., Hammen, V. L., Beukelman, D. R., & Traynor, C. D. (1990). The effect of rate control on the intelligibility and naturalness of dysarthric speech. The Journal of Speech and Hearing Disorders, 55(3), 550–560. https://doi.org/10.1044/jshd.5503.550