Impact of listeners perception and behavior

**What are the implications of finding certain human or synthetic voices to be more or less natural-sounding? What is the published evidence that naturalness affects behaviours in different contexts, for example human-human communication vs. human-computer interaction? Can you use these examples to convince the reader of the importance/timeliness of studying naturalness, as a basis for some of the more specific methodological criticisms and suggestions?**

Importantly, such variations in voice naturalness affect communicative quality [12,13]. Evidence from speech-language pathologies consistently shows that affected individuals with impairments in speech naturalness are perceived as withdrawn, cold, introverted or bored (Anand & Stepp, Klopfenstein, voijtech). This can result in social isolation, reduced quality of life and ultimately depression (Quellen). Importantly, these negative consequences even occur when intelligibility of the speaker is largely preserved (Quelle). Therefore, voice naturalness is a key target of speech therapy, across all types of voice alterations (Quellen). This is corroborated by a recent survey on personalized speech synthesis for people who lost their biological voice: almost two third of participants would prefer a more natural sounding voice, even at the cost of some degree of speech intelligibility, both as potential users as well as listeners (Hyppa-Martin 2024). Thus, for human-to-human interaction, there is no doubt that reduced voice naturalness has widespread negative implications. For human-machine interaction, the picture is less clear. Following the Computers-Are-Social-Actors (CASA) framework proposed in the 90s (Quelle), the assumption that we treat artificial agents like humans fueled an (implicit) naturalness-is-better bias for technical innovations (Quelle). Consequently, recent years have seen rapid developments in the effort to create synthetic voices that resemble human vocal expression [20,21]. Initial scientific findings suggest that this effort is justified, since reduced naturalness in synthetic voices was found to affect perceived likeability, trustworthiness, and pleasantness [11,22–25]. However, modern frameworks on synthetic voice design question a “one size fits all” idea and instead advocate solutions tailored to their specific application (Cambre). In that vein, maximum human-likeness of synthetic voices may not always be required nor desirable. Indeed, voice preferences for virtual agents seem to depend on features of the listeners (Eysell 2012; Lee 2010), the device itself (im 2023, McGinn and Torre 2019, Mitchell 2011) and its specific function (Im 2023, rodero 2017. Schreibelmayr 2022). Understanding and incorporating these preferences seems to be crucial for the success and acceptance of these devices (Lu 2021)

Importantly, such variations in voice naturalness affect communicative quality [Q]. Evidence from speech-language pathologies consistently shows that affected individuals with impairments in speech naturalness are perceived as withdrawn, cold, introverted or bored [Q]. This can result in social isolation, reduced quality of life and ultimately depression [Q]. Importantly, these negative consequences even occur when intelligibility of the speaker is largely preserved [Q]. Therefore, voice naturalness is a key target of speech therapy, across all types of voice alterations [Q]. This is corroborated by a recent survey on personalized speech synthesis for people who lost their biological voice: almost two third of participants would prefer a more natural sounding voice, even at the cost of some degree of speech intelligibility, both as potential users as well as listeners [Q]. Thus, for human-to-human interaction, there is no doubt that reduced voice naturalness has widespread negative implications. For human-machine interaction, the picture is less clear. Following the Computers-Are-Social-Actors (CASA) framework proposed in the 90s [Q], the assumption that we treat artificial agents like humans fueled an (implicit) naturalness-is-better bias for technical innovations [Q]. Consequently, recent years have seen rapid developments in the effort to create synthetic voices that resemble human vocal expression [Q]. Initial scientific findings suggest that this effort is justified, since reduced naturalness in synthetic voices was found to affect perceived likeability, trustworthiness, and pleasantness [Q]. However, modern frameworks on synthetic voice design question a “one size fits all” idea and instead advocate solutions tailored to their specific application [Q]. In that vein, maximum human-likeness of synthetic voices may not always be required nor desirable. Indeed, voice preferences for virtual agents seem to depend on features of the listeners [Q], the device itself [Q] and its specific function [Q]. Understanding and incorporating these preferences seems to be crucial for the success and acceptance of these devices [Q].