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pathologies consistently shows that affected individuals with impairments in speech naturalness can be perceived as withdrawn, cold, introverted or bored (Anand & Stepp, Klopfenstein, voijtech).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 system, 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 voice naturalness has widespread negative implications for both speakers and listeners. 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].

But now away from the “one size fits all” idea to tailored voice design which incorporated three layers: user, device and context (Cambre)

* User: Eysell 2012; Lee 2010
* Device: voice assistant embodied or disembodied (im 2023), people don’t like mismatch (McGinn and Torre 2019, Mitchell 2011)
* but the give the voice big importance
* Task, Objective: transport emotion and approachability – more humanlike is better (Ilves 2013)
* Context: specific to task (Im 2023, rodero 2017. Schreibelmayr 2022);
* difference between perception and interaction with these voices
* Ko 2023:
* nevertheless: the voice is an important feature in the design of virtual assistants and robots (Lu 2021)
* interferes with the processing of emotions (Duville 2022)