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?**

* People with a disorder that affects naturalness can lead to perception of these people sounding unhappy, cold, withdrawn, which leads to feelings of social isolation (in PD); Anand & Stepp 2015 and Voijtech 2019
* Disrupt communication, even when intelligibility is largely preserved
* Reduced naturalness can cause loss of independence and reduce quality of life
* Naturalness is an important outcome in speech therapy, for example for PD, stuttering (Frankford 2024, Euler 2021, Hardy 2020)
* Strong correlations between naturalness and acceptability (r\_= 90) in tracheoesophageal speech (Eadie 2008), tolerance for “unnatural voice features” is meditated heavily through the context and knowledge about the speaker (Eadie 2008, Goy 2016)
* Compelling case: voice banking and speech generative devices for ALS patients: almost two third of listeners indicated that they would prefer a more natural sounding system, even at the cost of some degree of speech intelligibility, for themselves and for others (Hyppa-Martin)

When interacting with human, naturalness is clearly a feature we are sensitive to, both as speakers and as listeners. For human-machine interaction, the story may be a bit more complicated. Fall in the naturalness-is-better bias.. (probably due to CASA) but that’s not necessarily the case.

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)