**Highlights**:

* Voices elicit impressions about their naturalness, which affect interactions between humans as well as with artificial agents
* Despite its intuitive appeal and practical importance, a systematic understanding of voice naturalness is elusive – the concept is scientifically ill-defined
* We show that current voice naturalness research is situated within different research domains that resemble echo chambers within science – they neither cross-refer to one another nor to current voice perception theory
* We offer a concise conceptual framework by proposing a taxonomy with two distinct types: deviation-based naturalness and human-likeness-based naturalness
* We develop practical recommendations and perspectives for naturalness research. We argue that, in a world of digital agents, understanding the determinants for how humans perceive naturalness in social stimuli is a priority