## *Vorschlag Sascha:*

## *Definitions of naturalness*

We propose a taxonomy with two distinct types: Deviation-based naturalness and human-likeness-based naturalness (**Figure 2**). In **deviation-based naturalness**, naturalness is defined as the deviation away from a reference that represents maximum naturalness. Example instructions for raters could be “Does this voice sound distorted?”, “Does this voice sound unusual?”, or just “Does this voice sound natural?”. This conceptualization needs two important specifications: the reference representing maximum naturalness, and the type of deviation. In some cases, the reference is explicitly provided e.g. through a comparison or baseline stimulus (see [78]). However, in many studies, raters are instructed to use an inner implicit reference that is based on their experience and expectations, e.g., judge whether “it conforms to the expected standard of unimpaired speech” [52]. The type of deviation is specified through the vocal material. It can virtually cover all acoustic features, ranging from specific manipulations (e.g., spectral features or speech rate [79–81]) to complex multivariate vocal patterns (e.g., in distorted or pathological voices [82]).

**Human-likeness-based naturalness** defines naturalness by its resemblance to a real human voice and is thus performed as an assessment of a reference. Instructions for raters could be “Does this voice sound like a real human speaker?” or “How human-like does the voice sound to you?” Compared to the deviation-based definition, the concept of human-likeness-based naturalness has an additional necessary and obligatory feature: the existence of a non-human voice space (e.g. artificially generated voice sounds) and occasionally a non-human sound space (voice-like sounds, e.g. some musical sounds or sound patterns). This highlights the notion of a categorical boundary to human voices, although the transition between categories can be continuous. In other words, a definition of human likeness is only essential if we assume that voices and voice-like sounds can be non-human in principle. Although the first type of deviation-based naturalness can, in certain cases, cross the boundary to the non-human voice space, this boundary is not essential for the definition of this type of naturalness. Apart from this critical distinction, human-likeness-based naturalness may represent a special case of deviation-based naturalness: the reference is a human voice (or listeners´ representation of a human voice), and the deviation is primarily assessed along the human/non-human spectrum. The above considerations especially suggest that the human-likeness-based conceptualization is particularly well-suited for research into synthetic voices.

## *Vorschlag Tine*

## *Definitions of naturalness*

We propose a taxonomy with two distinct types: Deviation-based naturalness and human-likeness-based naturalness (**Figure 2**). In **deviation-based naturalness**, naturalness is defined as the deviation from a reference that represents maximum naturalness. Example instructions for raters could be “Does this voice sound distorted?”, “Does this voice sound unusual?”, or just “Does this voice sound natural?”. This conceptualization needs two important specifications: the reference representing maximum naturalness, and the type of deviation. In some cases, the reference is explicitly provided e.g. through a comparison or baseline stimulus (see [78]). However, in many studies, raters are instructed to use an inner implicit reference that is based on their experience and expectations, e.g., judge whether “it conforms to the expected standard of unimpaired speech” [52]. The type of deviation is specified through the vocal material. It can virtually cover all acoustic features, ranging from specific manipulations (e.g., spectral features or speech rate [79–81]) to complex multivariate vocal patterns (e.g., in distorted or pathological voices [82]).

**Human-likeness-based naturalness** defines naturalness by its resemblance to a real human voice. Instructions for raters could be “Does this voice sound like a real human speaker?” or “How human-like does the voice sound to you?” Compared to the deviation-based definition, the concept of human-likeness-based naturalness has an additional necessary and obligatory feature: the existence of a non-human voice space. This highlights the notion of a categorical boundary to human voices, although the transition between categories can be continuous. In other words, a definition of human-likeness is only meaningful if we assume that voices can be non-human in principle. Although the first type - deviation-based naturalness - can, in certain cases, cross the boundary to the non-human voice space, this boundary is not essential for its definition. Apart from this critical distinction, however, human-likeness-based naturalness may represent a special case of deviation-based naturalness: the reference is a human voice (or listeners´ representation of a human voice), and the deviation is assessed along the human/non-human spectrum. The above considerations especially suggest that the human-likeness-based conceptualization is particularly well-suited for research into synthetic voices.