Abstract

Speech is made up of a continuous stream of speech sounds that is interrupted by pauses and breathing. As phoneticians are primarily interested in describing the segments of the speech stream, pauses and breathing are often neglected in phonetic studies, even though they are vital for speech. My PhD work adds to a more detailed view of both pausing and speech breathing with a special focus on the latter and the resulting breath noises, investigating their acoustic, physiological, and perceptual aspects. In my dissertation, I present an overview of how a selection of corpora annotate pauses and pause-internal particles, as well as a recording setup that can be used for further studies on speech breathing. For pauses, this work emphasized their optionality and variability under different tempos, as well as the temporal composition of silence and breath noise in breath pauses. For breath noises, we first focused on acoustic and physiological characteristics: We explored alignment between the onsets and offsets of audible breath noises with the start and end of expansion of both rib cage and abdomen. Further, we found similarities between speech breath noises and aspiration phases of /k/, as well as that breath noises may be produced with a more open and slightly more front place of articulation than realizations of 'schwa'. We found positive correlations between acoustic and physiological parameters, suggesting that when speakers inhale faster, the resulting breath noises were more intense and produced more anterior in the mouth. Inspecting the entire spectrum of speech breath noises, we showed relatively flat spectra and several weak peaks. These peaks largely overlapped with resonances reported for inhalations produced with a central vocal tract configuration. We used 3D-printed vocal tract models representing four vowels and four fricatives to simulate in- and exhalations by reversing air flow direction. We found the direction to not have a general effect for all models, but only for those with high-tongue configurations, as opposed to those that were more open. Then, we compared inhalations produced with the 'schwa'-model to human inhalations in an attempt to approach the vocal tract configuration in speech breathing. There were some similarities, however, several complexities of human speech breathing not captured in the models complicated comparisons. In two perception studies, we investigated how much information listeners could auditorily extract from breath noises. First, we tested categorizing different breath noises into six different types, based on airflow direction and airway usage, e.g. oral inhalation. Around two thirds of all answers were correct. Second, we investigated how well breath noises could be used to discriminate between speakers and to extract coarse information on speaker characteristics, such as age (old/young) and sex (female/male). We found that listeners were able to distinguish between two breath noises coming from the same or different speakers in around two thirds of all cases. Hearing one breath noise, classification of sex was successful in around 64%, while for age it was 50%, suggesting that sex was more perceivable than age in breath noises.