ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Acoustics of Breath Noises in Human Speech: Descriptive and Three-Dimensional Modeling Approaches

Raphael Werner,[a] Susanne Fuchs,[b] Jürgen Trouvain,[a] Steffen Kürbis,[c] Bernd Möbius,[a] and Peter Birkholz[c]

[a] Department of Language Science and Technology, Saarland University, Saarbrücken, Germany [b] Leibniz-Centre General Linguistics (ZAS), Berlin, Germany [c] Institute of Acoustics and Speech Communication, Technische Universität Dresden, Germany

ARTICLE INFO

Article History:
Received February 14, 2023
Revision received June 13, 2023
Accepted September 13, 2023

Editor-in-Chief: Cara E. Stepp
Editor: Hayo Terband

https://doi.org/10.1044/2023_JSLHR-23-00112

ABSTRACT

Purpose: Breathing is ubiquitous in speech production, crucial for structuring speech, and a potential diagnostic indicator for respiratory diseases. However, the acoustic characteristics of speech breathing remain underresearched. This work aims to characterize the spectral properties of human inhalation noises in a large speaker sample and explore their potential similarities with speech sounds. Speech sounds are mostly realized with egressive airflow. To account for this, we investigated the effect of airflow direction (inhalation vs. exhalation) on acoustic properties of certain vocal tract (VT) configurations.
Method: To characterize human inhalation, we describe spectra of breath noises produced by human speakers from two data sets comprising 34 female and 100 male participants. To investigate the effect of airflow direction, three-dimensional–printed VT models of a male and a female speaker with static VT configurations of four vowels and four fricatives were used. An airstream was directed through these VT configurations in both directions, and their spectral consequences were analyzed.
Results: For human inhalations, we found spectra with a decreasing slope and several weak peaks below 3 kHz. These peaks show moderate (female) to strong (male) overlap with resonances found for participants inhaling with a VT configuration of a central vowel. Results for the VT models suggest that airflow direction is crucial for spectral properties of sibilants, /ç/, and /i:/, but not the other sounds we investigated. Inhalation noise is most similar to /ə/ where airflow direction does not play a role.
Conclusions: Inhalation is realized on ingressive airflow, and inhalation noises have specific resonance properties that are most similar to /ə/ but occur without phonation. Airflow direction does not play a role in this specific VT configuration, but subglottal resonances may do. For future work, we suggest investigating the articulation of speech breathing and link it to current work on pause postures.
Supplemental Material: https://doi.org/10.23641/asha.24520585

Breathing, that is, repetitive cycles of inhalation and exhalation, is a vital activity of humans (and all aerobic organisms). In normal breathing at rest, inhalation and exhalation noises are present but often not audible to a typical listener. We can only speculate why this is the case: Inhalation and exhalation phases are more similar in duration than in speech production, and the glottis is widely open while the mouth is closed. In the context of speech production, breathing is indispensable: Exhalations are used for articulating speech and inhalations typically become audible. Research on the acoustics of breath noise, however, is relatively sparse, while practical applications may be manifold. Among others, the detection of breath noise can be revealing for the diagnosis of respiratory diseases, such as COVID-19 (Chen et al., 2022), and for pathological cries and coughs in infancy (Hirschberg, 1980). Breath noises can be affected not only by respiratory pathologies but also by changes in the vocal tract.

Correspondence to Raphael Werner: rwerner@lst.uni-saarland.de.
Publisher Note: This article is part of the Special Issue: Select Papers From the 8th International Conference on Speech Motor Control.
Disclosure: *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

Vocal fold paralysis due to injury of the recurrence nerve frequently results in a gap between the vocal folds. During speech production, this gap increases air consumption since expiratory air is constantly flowing through the glottis. The paralyzed vocal fold may also lead to a constriction in the vocal tract that can cause peculiarities in the temporal and spectral properties during inhalation. Spectral properties of inhalation noise vary according to speech task (Fuchs et al., 2023) and may also show traces of the participant's emotional (Goldman-Eisler, 1955) and cognitive state (Mitchell et al., 1996) during speaking.

In this article, we aim to close this research gap by focusing on breath noises produced in speech pauses with regard to three aspects: First, we will provide general (average) spectral descriptions of inhalation noises that are produced by a large number of human speakers. These spectral properties cannot be easily interpreted because the airflow direction in inhalation is ingressive while most spectra that researchers in phonetics are familiar with are realized with egressive airflow. Therefore, the second part of this article is dedicated to a better understanding of how spectra of inhalations and exhalations differ when keeping everything else but airflow direction constant. This was done using three-dimensional (3D)–printed vocal tract (VT) models. Finally, we compare human to model inhalations, aiming to approach the VT configuration of speakers when inhaling.

Inhalations are vital for speech, as they supply the lungs with the air that is needed to power speech production and occur every 3–4 s in most speech situations (Kuhlmann & Iwarsson, 2021; Rochet-Capellan & Fuchs, 2013; Winkworth et al., 1994), thus temporally structuring the flow of speech (Fuchs & Rochet-Capellan, 2021). Moreover, respiration is important for gas exchange with the environment and has an important biological function. During inhalation, oxygen is delivered to all parts of the body, and during exhalation, carbon dioxide is expelled (Hixon et al., 2020, pp. 38–39). Respiration structures brain activity (Heck et al., 2017), since the brain is among the largest oxygen consumers of the entire human body. Participants tend to inhale at the onset of cognitive tasks and also perform better when doing so, as opposed to exhaling (Perl et al., 2019). Such a phase-locking effect between inhalation and task onset was also found by Zöllner et al. (2021) for reaction time measures. While there is an extensive body of literature on speech breathing, the supraglottal mechanisms, as well as the resulting breath noises, have been largely neglected so far and thus remain underresearched.

Speech Breathing: Physiology and Acoustics

Switching from tidal breathing to speech typically reorganizes the breathing cycle: At rest, inhalations and exhalations are similar in duration, with inhalations being only a little shorter than exhalations (Conrad & Schönle, 1979; Werner, Trouvain, et al., 2021). When speaking, inhalations have a shorter duration and higher airflow velocity, while the exhalations, which are used for producing speech, become longer with a constant slow rate decrease in lung volume (Conrad & Schönle, 1979). At rest, inhalations take up around 40% of the duration of one breathing cycle, that is, one inhalation and exhalation (Gick et al., 2013, pp. 49–50), which in speech breathing is reduced to around 10% (Fuchs & Rochet-Capellan, 2021).

This reorganization also affects how air is inhaled: At rest, nasal inhalations are normative and deviating from that may even be detrimental to one's health (Hallani et al., 2008; Harari et al., 2010). Around speech, there is an additional demand such that inhalations have to take in enough air for gas exchange and to power speech breathing in a relatively short time so as not to interrupt the speech stream for too long (Conrad et al., 1983). This may be the reason that around speech, speakers usually deviate from the pattern of purely nasal inhalations that is prevalent at rest. Distinctions between different types of speech breathing, that is, nasal, oral, or alternations and combinations thereof, have been made in some studies: For instance, Kienast and Glitza (2003) and Scobbie et al. (2011) achieved a distinction of types by listening and/or looking at spectral characteristics. Lester and Hoit (2014) used nasal ram pressure to detect airflow through the nasal cavity. They found that a mixture of simultaneous oral and nasal breathing was used for the majority of speech breathing. However, nothing is known about moving articulators in the vocal tract other than the mouth opening. The general VT configuration in inhalations or more fine-grained aspects such as the degree of mouth opening or the behavior of the tongue remain largely unknown, as articulatory studies on speech inhalation are limited.

Related studies have looked at postures in speech initiation or pauses: Rasskazova et al. (2019) investigated the timing of acoustic, respiratory, and articulatory events before speech initiation using electromagnetic articulography and respiratory inductance plethysmography. They found speaker variability in the coordination of mouth opening and inhalation onset of thoracic volume change: Two speakers started inhaling before opening their mouths, one speaker first opened their mouth and then started inhaling, and the other three started both at roughly the same time. Others have investigated articulatory processes in speech pauses: Gick et al. (2004) compared articulatory settings in pauses of English and French speakers. They did find differences for most of their parameters but not for jaw aperture and velopharyngeal port width, even though, unlike English, the French phoneme inventory features nasal vowels.

They ascribed these similarities to physiological effects of inhalation that may be present in their pauses. Ramanarayanan et al. (2009) found differences in articulation, depending on whether the pause was planned at a phrase juncture or not. Krivokapić et al. (2022) looked at pause postures, that is, articulator movement occurring between the speech-related gestures before and after the pause that is not just an interpolation of the preceding and following gestures, in relation to speech planning. They found that longer upcoming utterances led to higher rates of occurrence of pause postures but not to longer pause durations. What these studies on articulation in speech pauses or before speech typically have in common is that they do not differentiate between acoustic-phonetically silent nonbreath pauses and breath pauses. Therefore, it can only be speculated, as is done by Gick et al. (2004), whether and to what degree the results have been affected by speech breathing.

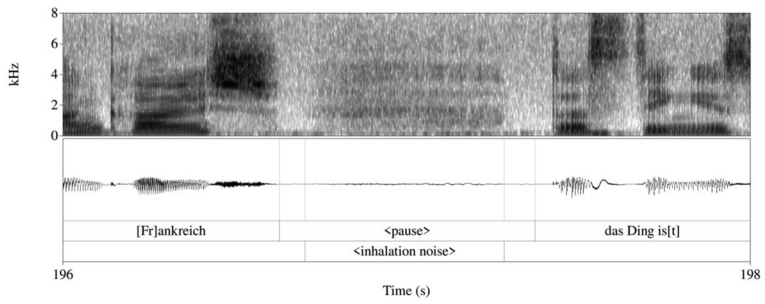
Figure 1 shows a typical example of an inhalation noise embedded in a speech pause, surrounded by short edges of silence around them (Ruinskiy & Lavner, 2007). It has a weak formant-like structure, as well as noise in the frequency range of up to 4–5 kHz. Few studies have looked at the acoustics of breath noises: Nakano et al. (2008) investigated them in singing with the aim to improve breath detection algorithms. With regard to their acoustic characteristics, they found breath noises to have similar spectral envelopes within the same song and also within the same singer. They also note that the long-term average spectra of breath noises found in male singers have a peak at 1.6 kHz and those of female singers at 1.7 kHz. Along with such spectral peaks in the second formant (F2) region, they found secondary peaks that exist in the range of 850–1000 Hz, which are, however, more prominent in female singers.

Werner, Fuchs, et al. (2021) used center of gravity (CoG), intensity, and the first three formants (F1–F3) to compare inhalations to speech sounds like the voiceless glottal fricative [h] and [ə], as well as aspiration phases of [p^h t^h k^h]. They found CoG values of inhalations to be similar to the aspiration of [k^h] but not [p^h t^h]. For formants, inhalations tended to have higher F2 values but particularly higher F1 values than [ə], suggesting a more open and slightly more front articulatory posture for inhalations. CoG, intensity, and F1 were also positively correlated with inhalation slope, that is, when speakers inhaled more air in a given time, the resulting breath noises had a higher CoG, intensity, and F1 values. Like Nakano et al. (2008), they noted F2 to be typically more prominent than F1. A limitation of this study, and of those using speech sounds as references in general, is that inhalations are typically produced without phonation and with an air stream direction that is opposed to nearly all speech sounds. In this sense, formant values may not be comparable, because the airflow direction may affect the spectral characteristics of the respective sounds.

Some other attempts to characterize the acoustics of breath noises include the analysis of their formants in snoring: Ng et al. (2008) found F1 to be much lower for benign versus apneic snorers (360 vs. 724 Hz on average), while F2 (1627 vs. 1809 Hz) and F3 (2840 vs. 2955 Hz) did not differ as much.

When analyzing breath noises, there may also be an influence of subglottal resonances that is stronger than in speech sounds. The degree of coupling between the supra- and subglottal tracts depends on the glottal state (Lulich, 2010): With an infinite glottal impedance, that is, with a closed glottis, there is no coupling between the two tracts, and the resulting poles are the natural frequencies of the supraglottal tract. With a partially open glottis, there is some coupling, which leads to the VT poles being shifted upward in frequency and new poles being introduced from the subglottal tract. This increase in frequency was also demonstrated experimentally for phonation in VT models with an increasing peak glottal area (Birkholz et al., 2019). In breathing, and inhalations especially, however, the glottis is opened much wider, with male adults reaching a peak glottal area of 217 ± 54 mm^2 ($M \pm SD$) during slow and 228 ± 43 mm^2 during rapid inhalations

Figure 1. Spectrogram of a section from the Pool 2010 corpus containing an inhalation noise located in a speech pause surrounded by silent edges.



(Scheinherr et al., 2015). For females, the peak glottal area is 189 ± 32 mm^2 for slow breathing and 184 ± 25 mm^2 for fast breathing.

Hanna et al. (2018) examined how the impedance spectrum measured through the lips is affected by the subglottal tract in varying states of glottal opening, that is, fully closed, fully open, as well as intermediate states for phonation and respiration. In their experimental part, they used impedance spectrometry to measure resonances and antiresonances in 10 Australian participants who were instructed to keep their tongue in the position of the /ɜː/ vowel. This vowel in Australian English should be very close to the VT configuration of /ə/, with the former being slightly more open. Participants were also instructed to keep their velum raised. As the velum may be hard to raise volitionally, some participants in Hanna et al. (2016) are reported to have pinched their noses to avoid nasal participation. For inspiration in the seven male participants, Hanna et al. (2018) found impedance minima, which correspond to resonances and can be seen in Table 1. They showed that combining the supra- and subglottal tract, as is done in respiration, effectively doubled the length of the tract, which leads to up to twice as many resonances and antiresonances compared to the closed glottis condition.

In summary, the acoustic characteristics of breath noises have been examined very rarely in general. Most of the few pertinent studies investigated the acoustics of breath noises in snoring or singing. While snoring is a specific subtype of breathing, influenced by the supine position and unrelated to speech production, trained singers may also train their breathing and, thus, may differ from the rest of the population (Salomoni et al., 2016). Spectral descriptions of breath noises in speech, however, are rare and often based on a few participants.

Change of Airflow Direction: Inhalations Versus Exhalations

One specific property that has an effect on the spectral characteristics of breath noise and its interpretation is airflow direction. The majority of speech sounds are produced with a pulmonic egressive air stream, while ingressive pulmonic phonation is far less prevalent (Eklund, 2008). This may be related to the fact that vocal fold anatomy is better suited for phonation with a pulmonic egressive rather than ingressive air stream (Catford, 1977, pp. 67–68).

Catford (2001, pp. 18–21) described that the two voiceless fricatives [f] and [s] can also be produced with pulmonic suction, that is, while inhaling (as expressed by the International Phonetic Alphabet symbol [↓]) rather than exhaling. The inhalation would then affect the sounds [f↓] and [s↓] differently: [f] and [f↓] would not be very different, as the channel through which air flows is not changed much as a function of direction; hence, the turbulent airflow and the resulting sound are similar. For [s] and [s↓], however, the egressive version flows through the narrow gap between the tongue and the alveolar ridge and leaves that channel in a high-velocity turbulent jet. This jet then hits the teeth, making it even more turbulent, which adds an additional high-frequency component to the sound of [s]. The ingressive version [s↓] involves comparably slow, nonturbulent airflow past the teeth before it reaches the narrow channel between the tongue and the alveolar ridge. There, it is accelerated and becomes turbulent. In this case, however, there is no obstacle present, like the teeth in the egressive version, so there is no added turbulence, and thus the resulting sounds differ.

Our ultimate goal in this article is to estimate the potential VT contributions to the spectral characteristics of inhalation noise. Doing so requires not only a description of the acoustic properties but also an analysis of how airflow direction could affect it. A better understanding of breath noise, airflow direction, and the potentially underlying VT contributions is not only important for the recognition of breath noise in speech corpora but also an important baseline for clinical applications.

Inhalations in Human Speakers

Method

Material

We used data from two sources. One is the Pool 2010 corpus (Jessen et al., 2005), from which we used semispontaneous speech produced by 100 male native speakers of German (M_{age} = 39 years, range: 21–63). Each speaker was recorded in two conditions: a Lombard condition, in which they heard white noise via headphones while speaking, and a "normal," non-Lombard condition. The data used here were taken from the non-Lombard speech condition. For this task, speech was elicited in a setup similar to the game Taboo, that is, speakers were

Table 1. Impedance minima corresponding to resonances (in Hz; mean and standard deviation) reported for seven male and three female participants inhaling with a vocal tract configuration of /ɜː/ in Tables I and II in Hanna et al. (2018).

Male	530 (60)	880 (55)	1335 (145)	1735 (70)	2210 (75)	2565 (95)	3660 (270)
Female	660 (30)	1020 (45)	1490 (45)	1820 (65)	2460 (55)	2790 (160)	3680 (455)

asked to describe pictures to a conversation partner without using several terms. The recordings were made with a sampling rate of 16000 Hz.

The other source consists of 34 female native speakers of German (M_{age} = 25 years, range: 20–33) from the data set described in Rochet-Capellan and Fuchs et al. (2013). The participants produced semispontaneous speech, each retelling five short stories (fables). The sampling rate in these recordings was 11025 Hz. For the data set of female speakers, we only used frequencies up to 4500 Hz and thus downsampled it to a sampling rate of 9000 Hz for two reasons: The audio files had shown a strong intensity decline in the higher frequency regions before, and downsampling to a sampling rate divisible by 50 allowed us to have the same spectral resolution of 50 Hz across data sets for better comparability (see the Data Analysis section for details).

We used on- and offset of noise in the audio signal to annotate inhalations in both data sets. This resulted in 1,892 inhalations produced by male speakers in the first source and 749 inhalations from female speakers in the second source. Both sources included stretches of participant inactivity in the audio files, for instance, when the speaker was quiet between tasks. While speakers must breathe between these tasks, breath noises there were hardly visible in the spectrogram. We only used inhalations occurring in speech pauses, that is, those preceded and followed by speech, because we focused on speech inhalations, while breathing at rest may differ. This also excludes breath noises that may be related to turn-taking. All 2,641 inhalations produced by human speakers were extracted via a Praat script (Boersma & Weenink, 2019) using rectangular windows. On average, the 134 speakers contributed a mean number of 19.7 inhalations (SD = 13.2, range: 1–61). The mean duration of these breath noises was 467 ms (SD = 225 ms) for the male speakers and 410 ms (SD = 147 ms) for the female speakers.

Data Analysis

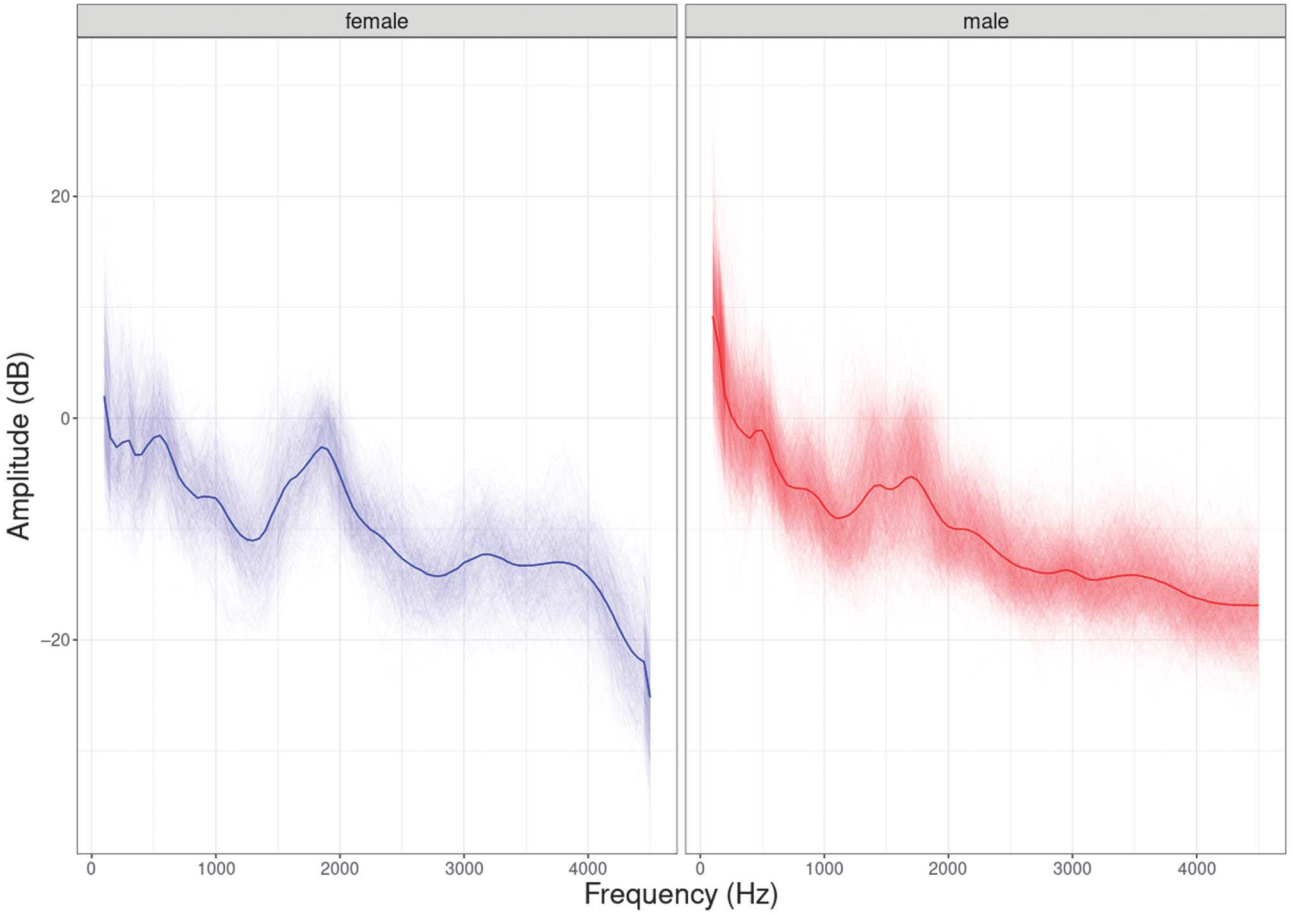
Although other studies have used formants to describe breath noises (Nakano et al., 2008; Werner, Fuchs, et al., 2021), here we used the averaged power spectral density (PSD) of the sounds, because formants are inherently difficult to determine in voiceless speech and breath noises and not uniquely defined due to the spectral zeroes introduced by coupling with the subglottal system. In addition, formant values of unvoiced vowels, for instance, in whispered speech, seem to constantly deviate from voiced vowels (Heeren, 2015), thus complicating interpretation. Using PSD also gives us the advantage of analyzing the entire spectrum, rather than extracting just one value. We obtained PSDs via pwelch in MATLAB using a Hamming window. For the male speakers, recorded with a sampling rate of 16000 Hz, we used a window length of 320 points with a 50% overlap between windows. This results in a spectral intensity value every 50 Hz from 0 to 8000 Hz. For the female data, we followed the same procedure but adjusted the window length to 180. With the data downsampled to 9000 Hz, we also have a spectral resolution of 50 Hz here to make the data sets comparable. We removed the measurement points at 0 and 50 Hz, as these low-frequency components tend to be problematic and differences there could arise from different recording setups.

All human inhalation spectra in both data sets underwent the same preprocessing steps, so they had the same frequency range and spectral resolution, and the amplitudes of the spectra in both data sets were normalized together: To normalize, we obtained the sum of spectral components for every spectrum and divided it by the number of components, that is, the number of spectral intensity values per spectrum. We chose the median of these sums as the arbitrary target sum, which we then subtracted from the sum of every individual spectrum to compute the difference between the target sum and the sum of an individual spectrum. As a final step, we subtracted this difference from every component. This allowed us to maintain the shape of each spectrum while shifting them in amplitude toward the same sum of components. Using data from two sources and a larger number of speakers makes this even more necessary, as recording setups are likely to vary. For the normalization, it was necessary to have the same frequency range for male and female data, namely, 100–4500 Hz. The unnormalized inhalation spectra averaged by sex are shown in Supplemental Material S1. It can be seen that for male speakers, the spectrum hardly has any peaks above 4500 Hz. The inhalation spectra averaged by speaker can be seen in Supplemental Materials S2 and S3. While there is some by-speaker variation, inhalations show some common patterns between speakers.

Results and Discussion

Figure 2 shows all the human inhalation spectra and their average split by sex. Both of them are relatively flat in comparison to speech sounds with a decreasing slope from higher intensity for low frequencies to lower intensity for higher frequencies. They do have several weak peaks: seven in the male data and eight in the female data. The strongest of these can be seen below 2 kHz; it is highest at around 1.85 kHz for women and 1.7 kHz for men. In both cases, there is also a slightly weaker peak a little below, at around 1.6 kHz for females and 1.45 kHz for males. Both speaker groups also show a peak in the region of 500 Hz, highest for female speakers at around 550 Hz

Figure 2. All human inhalation spectra for female (blue) and male (red) data. The average spectrum per sex is overlaid in bold.

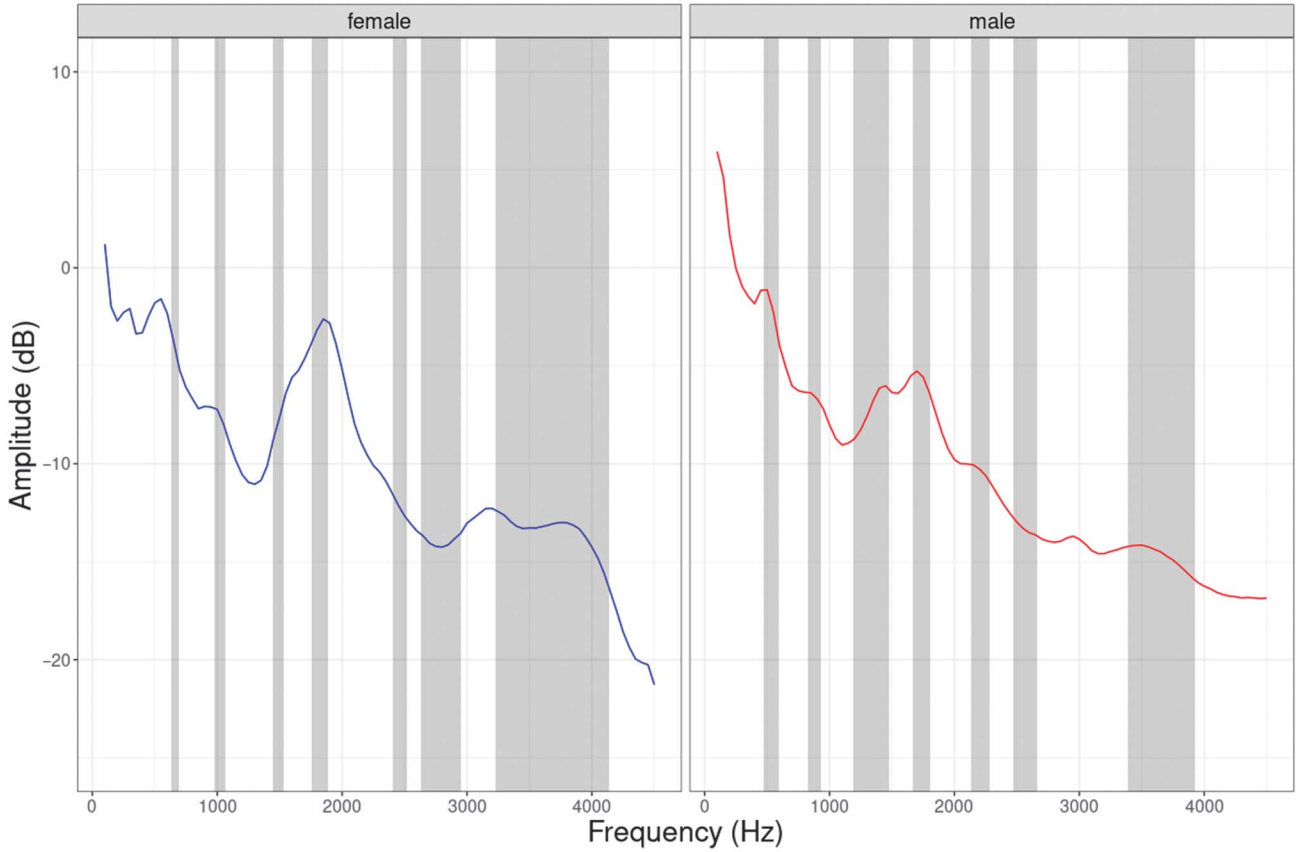


and for male speakers at around 500 Hz. In addition, both spectra have very weak peaks at around 1, 2.2, 3, and 3.5–4 kHz. Only the female speakers have an additional peak at 300 Hz, which is absent from the male speakers. Moreover, there is not a lot of variation between individual inhalations, as they are all very similar within one sex. We suspect the differences in the spectra by sex to be caused by differences in average body height between the two speaker groups, as this is generally related to airway size. When normalizing for differences in these volumes and tract lengths, speech breathing is generally the same for men and women (Hixon et al., 2020, p. 57).

The spectra shown here are similar to those reported by Nakano et al. (2008), who found spectral peaks at about 1.6 kHz for males and 1.7 kHz for females, although both are slightly higher in our data. The secondary peak, which they found to be stronger for female than for male speakers at around 850–1000 Hz, occurs at lower frequencies, that is, around 500 Hz, in our data. Similar to their findings, it is slightly stronger for female speakers.

Figure 3 shows the average spectra by sex in relation to the resonances reported for male and female speakers inhaling with a VT configuration of /ɜː/ in Hanna et al. (2018). It is striking that for male speakers, all of the resonances they reported, except for the sixth, also align with peaks in our data. For female speakers, the data do not overlap as clearly but are still similar. The sixth resonance, as for male speakers, should also be higher according to our data. However, most of the shaded resonances are close to peaks in our spectra. The peak at 300 Hz, which only showed up in our female speakers, is absent from their findings. A possible reason for why the resonances align better for male speakers may be that both our study and Hanna et al. (2018) have more male speakers (100 and seven, respectively) than female speakers (34 and three, respectively). In addition, our data were elicited under more natural conditions than theirs, which were recorded in a very controlled setting, for which participants were asked to keep their velum, VT configuration, and glottal opening constant for several seconds. Our data are likely to include more

Figure 3. Averaged human inhalation spectra for female (blue) and male (red) data. The regions shaded in gray indicate resonance bands for inhalations with the vocal tract configuration of the vowel /ɜ:/, as reported for male speakers in Table I and female speakers in Table II by Hanna et al. (2018).



articulatory movements, as well as some inhalations happening partly or entirely through the nasal tract. It should be mentioned that male and female speakers here also differ by speech task, that is, Taboo game versus retelling a story that may involve different cognitive load, and by age (M_{age} = 39 vs. 25 years). In summary, our human inhalation spectra suggest that the male speakers and, to some degree, the female speakers may inhale similarly to the inhalation condition in Hanna et al. (2018), that is, with a central VT configuration and a widely opened glottis, coupling the supra- and subglottal tracts.

Inhalation Versus Exhalation in 3D-Printed VT Models

In this part, we focused on the effect that a reversion of the air stream direction, that is, ingressive versus egressive, has on the acoustic characteristics of the resulting noise. For this purpose, we used 3D-printed VT models to study the effect of direction in the exact same VT configuration.

Method

Material

We used 3D-printed VT models (see Figure 4), as described in more detail in Birkholz et al. (2020). For these, a male (age: 39 years, height: 1.85 m) and a female (age: 32 years, height: 1.64 m) native German speaker were asked to produce several sustained speech sounds while capturing a volumetric magnetic resonance imaging (MRI) scan of their VTs. Maxilla and mandible shapes were included via plaster models. The models we used did not include nasal cavities. The VT models were 3D-printed using polylactic acid, a commonly used filament material. This results in the VT walls being hard in comparison to the soft VT walls in humans.

We here use a subset of these VT configurations, representing several sounds, namely, four vowels /i: a: u: ə/ and four fricatives /x ç ʃ s/. We chose /ə/ because we

Figure 4. Two of the 3D-printed vocal tracts corresponding to a male speaker producing the sounds /a:/ (left) and /ʃ/ (right). 3D = three-dimensional.

assume a similar configuration for inhalations. The peripheral vowels /i: a: u:/ are used as reference, and the fricatives were chosen because breath noises typically have fricative-like acoustics (Székely et al., 2019). To imitate inhalations and exhalations, the VT models were supplied with static airflow through a constantly open glottis (diameter: 10 mm; glottal area: 78.54 mm^2) at three fluid power levels in two airflow directions. The power levels were 500, 1,000, and 2,000 mW and were chosen to roughly simulate quiet breathing, loud breathing, as well as an intermediate level. The open glottis was connected to the artificial lung via a polylactic acid trachea of 20 cm in length (diameter: 17 mm) and a bronchial horn of 7 cm in length. Below the glottis, the trachea is tapered from 17 to 10 mm over a length of 30 mm. The generated noises were recorded for 10 s each with a microphone with a sampling rate of 48000 Hz. Overall, this results in 96 audio recordings of modeled breath noises (8 vocal tract configurations × 2 directions × 2 model speakers × 3 power levels).

Data Analysis

To obtain the spectra, we used similar methods as described in the Method section of Inhalations in Human Speakers. Given the sampling rate of 48000 Hz, we set the window length in pwelch to 960 samples with a 50% overlap between windows, which resulted in a spectral resolution of 50 Hz. For the analysis, we chose to filter out all measurement points greater than 10 kHz. We also removed the measurement points at 0 and 50 Hz as in the Data Analysis section above.

We calculated the discrete cosine transform (DCT) coefficients 0–3 to further characterize and compare the sound spectra via the RStudio (RStudio Team, 2022) package and function emuR::DCT (Winkelmann et al., 2021). The DCT models a function, that is, the magnitude spectrum in our case, as a weighted sum of orthogonal cosine functions with increasing frequencies. In this sense, the DCT is similar to the Fourier transform but uses a different set of basis functions. The DCT coefficients are the weights of the cosine-shaped basis functions, and the higher the DCT index, the finer the spectral detail it represents. From another point of view, the DCT can be seen as a way to decorrelate the points of discrete signals (Ahmed et al., 1974). DCT0 has been used as corresponding to a spectrum's mean amplitude, and DCT1 has been used as corresponding to its slope (Jannedy & Weirich, 2017). We then fitted a separate linear mixed-effects model for each DCT coefficient with direction (two levels: inhalation vs. exhalation) and vocal tract configuration (eight levels), as well as their interactions, as predictors. The models also included random intercepts for speaker and power level. We used lme4 (Bates et al., 2015) for model fitting and emmeans (Lenth, 2021) for pairwise post hoc comparisons between inhalations and exhalations for each configuration. To correct for multiple comparisons, we used the standard procedure in emmeans, adjusting p values via the Tukey method for comparing a family of 16 estimates for each of the statistical models. All models had at least one significant interaction between direction and configuration except the one for DCT3, so we used an additive model for DCT3. To avoid singular fit warnings, for DCT2, we used the linear model lm(DCT2 ~ direction * VTconfig) without random effects. For DCT0 and DCT1, we used the following model formulae: lmer(DCTi ~ direction * VTconfig + (1|speaker) + (1|condition)), with i being 0 or 1. In the case of DCT3, * was replaced by + and i was 3.

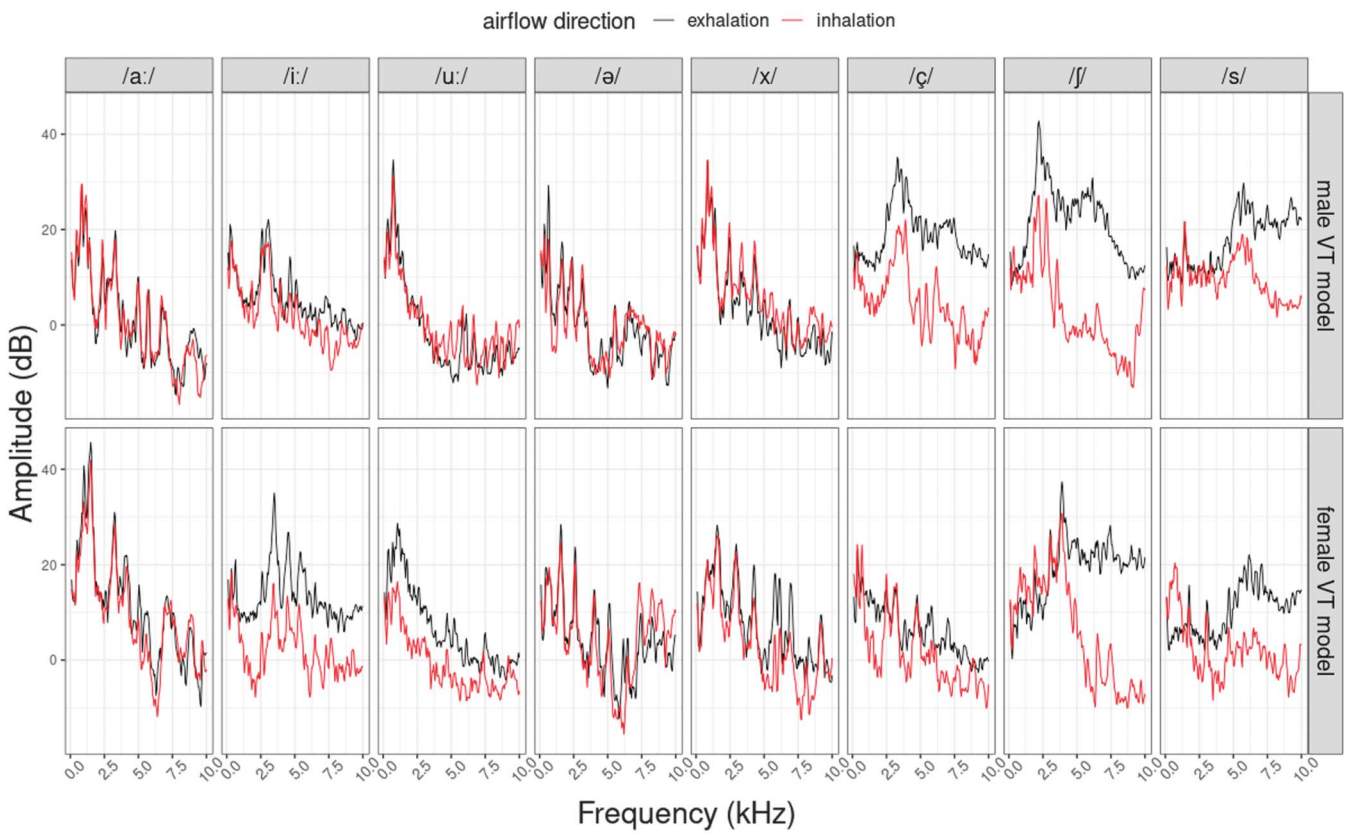
Results and Discussion

Airflow Direction: Inhalation Versus Exhalation in VT Models

The resulting spectra can be seen in Figure 5. For every combination of VT model, VT configuration, and direction, the three power levels are averaged for better readability, as they mainly differ in amplitude. The plot shows that for some VT configurations, a change in airflow direction entails stronger spectral differences than for others.

The statistical analysis revealed no general effect of reversing airflow direction on the spectrum in any of the four statistical models for DCT0–3. For each of the four models, there were several main effects for VT configurations that were, however, not of interest to our question. Importantly, we found interactions between airflow direction and VT configurations, that is, differences between inhalation and exhalation that were specific to some VT configurations. The statistical output for the pairwise post hoc comparisons between directions that were significant (after Tukey adjustment) using an α level of .05 can be found in Table 2. DCT0 was significantly higher for /i: ç ʃ s/

Figure 5. Spectra (100–10000 Hz) for exhalation (black) and inhalation (red) by vocal tract (VT) configuration and VT model (male and female). Every spectrum averages over three power levels.



with egressive flow. DCT1 values were significantly higher in ingressive flow for /ʃ s/. For DCT2, the results from the linear model suggested that /ʃ/ was the only VT configuration that showed an effect of airflow direction, as it was higher in ingressive airflow. For DCT3, there were no significant interactions between direction and VT configuration; thus, we do not have any significant contrasts between directions.

The direction differences were thus mostly found in sibilants and /ʃ/ especially. For the mean amplitude, as expressed by DCT0, we found differences for four VT configurations, all of which featured a high tongue position. Here, we assumed that the tongue height led to a concentrated airstream hitting the incisors. While this obstacle source amplified the signal in exhalation, with a reversed airstream, there was no concentrated airstream hitting the incisors, which is why the signal was much weaker in inhalations. /s/, but also /ʃ i: ç/, thus following Catford's (2001) prediction, as they showed high-frequency components in exhalation, which they did not have in inhalation.

Table 2. Significant contrasts between airflow direction by vocal tract (VT) configuration in the VT models (exhalation–inhalation).

Coefficient	VT configuration	Est.	*SE*	*df*	*t* ratio	*p* value
DCT0	/i:/	10.58	2.95	77	3.58	.0444
	/ç/	13.50	2.95	77	4.58	.0018
	/ʃ/	22.42	2.95	77	7.60	< .0001
	/s/	11.95	2.95	77	4.05	.0108
DCT1	/ʃ/	−12.44	1.12	77	−11.12	< .0001
	/s/	−9.08	1.12	77	−8.11	< .0001
DCT2	/ʃ/	−5.74	1.12	80	−5.12	.0002

Note. DCT = discrete cosine transform.

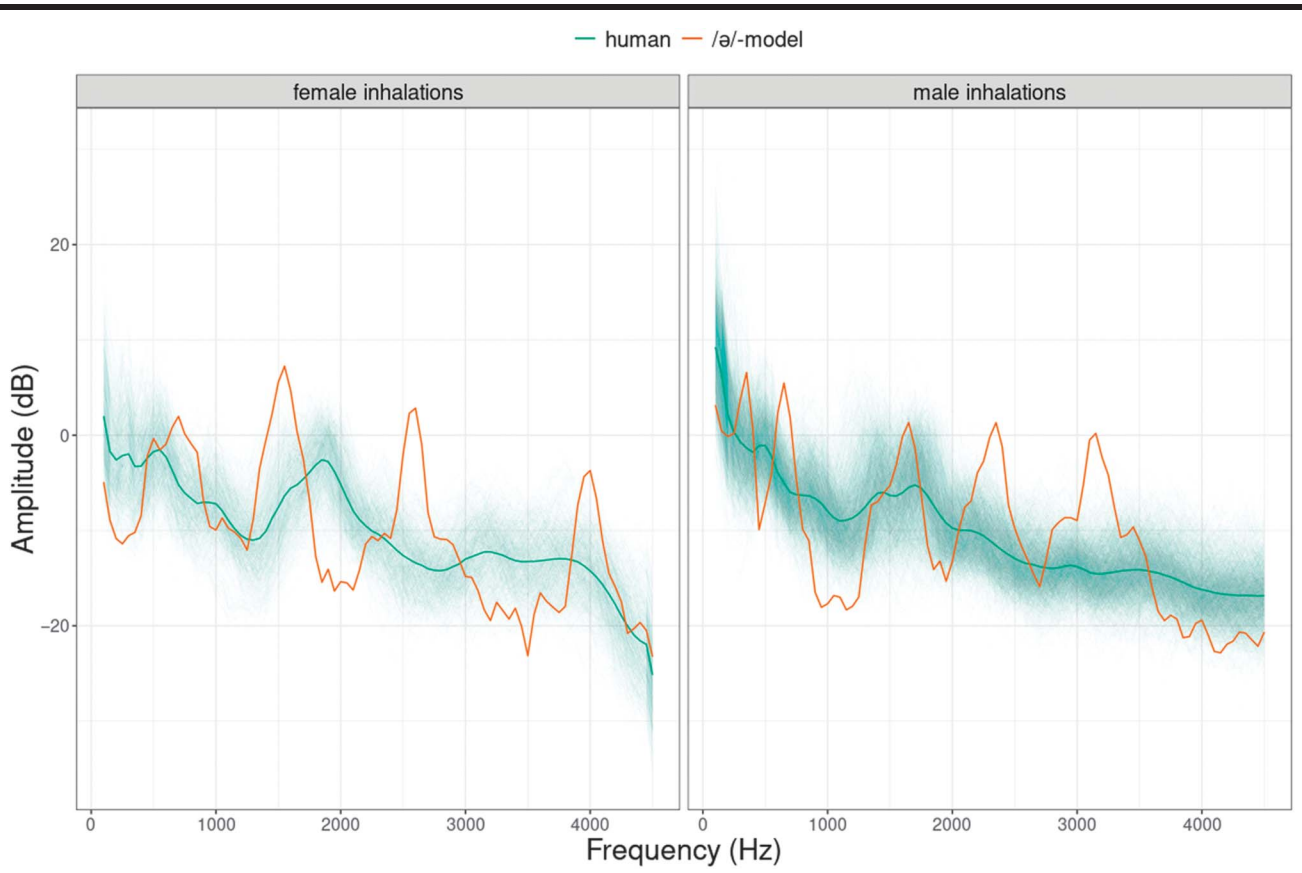
It should be mentioned that the 3D-printed VTs are based on MRIs of a single male and a single female speaker, respectively. Therefore, we could not determine if differences between the two with regard to a change in airflow direction were based on sex or idiosyncratic differences. This can be seen to varying degrees for /i: u: ç ʃ/.

VT Configuration: Human Versus Model Inhalations

We here compare human inhalations with model inhalations produced with an underlying /ə/-VT configuration. This builds on our previous findings on airflow direction in unconstricted VTs and on the similarity of human inhalation to experimental findings on inhalations with the VT of a central vowel. For this, we used the results from human inhalations from the Inhalations in Human Speakers section and the /ə/ inhalations from the models, thus excluding exhalations. We only included data up to 4.5 kHz, as this was the maximum for the female human data. Afterward, we removed the first two measurement points at 0 and 50 Hz and normalized the amplitude within each of the two data sets as described in the Data Analysis section. Rather than using quantitative methods, we chose to employ a qualitative approach to compare spectra instead, as several differences between human and model VTs complicate the comparison and need to be taken into account.

The comparison between human inhalation and model inhalation with a /ə/ VT can be seen in Figure 6. In the /ə/ models, there are stronger but also fewer peaks compared to human inhalations, with four major peaks for the female model and five for the male model in the frequency range up to 4.5 kHz. There are some similarities regarding their locations. The two main peaks in human inhalations, that is, around 500 Hz and below 2 kHz, have corresponding peaks in the respective model data. In the male data, this is visible for the higher of the two especially, whereas the model has two peaks below and above 500 Hz. In the female data, the two peaks are closer to each other, that is, the lower one is shifted upward in frequency while the higher one is close to 1.5 kHz. The human inhalations only have weaker peaks beyond 2 kHz, while for the model /ə/, the peaks remain almost as or equally strong in higher frequencies.

Figure 6. Human inhalation spectra (green; all spectra and average) versus model inhalation with a /ə/ vocal tract configuration (orange; averaged over three power levels each). The data are split by sex.



A possible reason for why model /ə/ and human inhalations align better in the male data may be, besides the difference in sample size of human inhalations, that the synthetic subglottal tract and glottis that were used in our setup were the same throughout the recordings and only the supraglottal VT was changed for male and female models. In human speakers, the length of the subglottal tract differs by sex, with 19.5 cm for male and 16.0 cm for female participants on average (Hanna et al., 2018). In human participants, the size of the glottal opening in breathing also differs by sex (Scheinherr et al., 2015).

In addition to that, there are several ways in which inhalations in the VT models differ from human inhalations, which might complicate the comparison based on acoustics and led to our decision of comparing them qualitatively. First of all, the VT models do not have a nasal tract. Following Lester and Hoit (2014), the majority of human speech inhalations we have in our data set may be simultaneously nasal and oral, which could affect the spectral properties by enlarging the VT's surface area and volume. In addition to that, there may also be purely nasal inhalations or alternations of oral and nasal airway usage. Another level of complexity not captured in the models is vertical larynx movement (Fink, 1974; Orlikoff et al., 1997). In human inhalation, the larynx moves downward, lowering the trachea, stretching out the laryngeal soft tissues, and flattening the vocal folds against the side wall. The displacements are then reversed in exhalation. With the models being static, they do not account for this vertical larynx movement modifying the lengths of the supra- and subglottal cavities, which may affect the spectral properties of inhalations. The lack of movement in the models may also differ from human inhalation, where speakers are unlikely to hold the exact same configuration in their vocal tract throughout the inhalation, as opening and closing gestures of the mouth and the velopharyngeal port may fall into that phase, as well as potentially coarticulatory adaptation to surrounding gestures. Moreover, the VT models were printed using hard plastic, which is different from the soft, fleshy human VT. Although these differences are not substantial, using soft VT models may have increased the frequency and bandwidths of the first resonance and partly of the second resonance, and the effect is quite vowel dependent (Birkholz et al., 2022). The resonators with soft walls in the referenced study were simple axisymmetric tubes, made with molds and silicone. As far as we know, flexible filament available for 3D printers is too stiff to be suitable as material for soft vocal tract walls. For more complicated MRI-based vocal tract shapes, variants with soft walls have not been made so far, but we would expect that the results are the same as for the axisymmetric tubes. Additionally, the glottal area is likely to be different in human and model inhalation. While in human inhalations, it would differ between fast and slow breathing (see the Speech Breathing: Physiology and Acoustics section) and by airflow direction (217 mm^2 for inhalation and 178 mm^2 for exhalation of male speakers according to Scheinherr et al., 2015), in the models it was constant across power levels and airflow directions (see the Material section of Inhalation Versus Exhalation in 3D-Printed VT Models). However, since changes of small cross-sectional areas affect the resonance frequencies much stronger than changes of larger areas, the precise glottal area is not so decisive. Furthermore, the glottis can be regarded as a relatively short section of the airways (3–5 mm). According to the theory outlined by Ungeheuer (1969), the cross-sectional changes of very short sections of the vocal tract mainly affect resonances at higher frequencies. Hence, overall, the exact value of the glottal area (as long as it is not too small) has little effect on the frequencies of the first few resonances. Finally, the VT models are based on the anatomy of a single speaker per sex. Even if these two speakers represented the respective average VT, we would still expect a substantial degree of variation between speakers. We used these speakers as prototypical starting points, similar to tongue and vocal tract models that are based on one speaker, for example, VocalTractLab (Birkholz, 2013), Gepetto (Patri et al., 2015), or the Dang model (Dang & Honda, 2002).

General Discussion

In the first part of this article, we characterized male and female human inhalations acoustically, which were quite similar within a sex group. Their spectra were relatively flat but showed several weaker peaks that aligned with resonances for inhalation with a VT setting of a central vowel reported by Hanna et al. (2018).

In the second part of this article, we found that reversing airflow direction does not yield a general effect on the resulting audio signal. We did, however, find an effect on the mean amplitude of the spectrum for VT configurations with a high tongue position, as well as on the slope for both sibilants and DCT2 for /ʃ/.

Following the findings from the first two parts, we compared the spectra of human inhalations and model inhalations produced with a /ə/ VT configuration. There, we found that the human inhalation spectra show similarity to modeled /ə/ inhalations. However, many aspects of the physiology of speech breathing are underresearched, which makes it difficult to neatly tease all the factors apart that potentially contribute to the spectral properties.

Therefore, more articulatory studies with a focus on speech breathing are necessary to learn in detail about the VT configuration speakers adopt in speech inhalations.

There are numerous influences on the acoustics beyond the supraglottal configuration, including the subglottal tract, glottal opening, nasal participation, and larynx lowering. Since the velum and the larynx are difficult, or even impossible, to track via electromagnetic articulography, real-time MRI is arguably the best method to do this. Modeling inhalations by means of 3D-printed VT models help us learn more about how speech breathing is performed. However, there are several complexities and aspects of temporal coordination the models cannot capture.

In this study, we chose to use the power spectral densities of breath noises to investigate their spectral properties and thus decided against using parameters such as formants or CoG. The reasoning behind this was that, given the scarcity of studies on the spectral properties of breath noises in comparison to speech, we wanted to examine the spectrum as a whole, rather than extracting single values from it. In addition, we were not convinced that they would be a good fit for our research, as formants are difficult to determine in the absence of phonation. CoG is most informative for those fricatives that have strong concentrations of energy in certain frequency regions, which was not the case for the inhalation spectra investigated here. With the description of inhalation spectra presented here, future studies could explore which other parameters work best to describe inhalation noises.

Conclusions

In this study, we examined the spectral properties of speech inhalations with the goals to describe human inhalations, to investigate the effect of reversing airflow direction in inhalations and exhalations in VT models, and to approach the VT configuration in human inhalations by comparing them to modeled inhalations produced using VT models whose underlying configurations we know. We found that human inhalations have relatively flat spectra with decreasing slope and several weak peaks. The peaks showed moderate (female data) to strong (male data) overlap with resonances found for participants inhaling with the VT configuration of a central vowel by Hanna et al. (2018) that arise due to coupling of subglottal and supraglottal tracts. For the VT models, results suggested that airflow direction has a segment-specific rather than a general effect on the acoustic properties, which affected especially the realizations of /ç/, /iː/, and sibilants. Hence, the spectra of inhalation noises, produced with ingressive airflow, can be compared with speech sounds involving egressive airflow, given that they do not have a small tongue constriction. We further found similarities between human inhalations and modeled inhalations with the VT configuration of /ə/. However, several differences between the model and human inhalation, as well as aspects of the physiology of speech breathing that are yet to be researched, complicated the comparison. This study is a first step toward modeling breath noises. Future attempts should incorporate further complexities of human inhalation, such as soft tissue, laryngeal excursion, dynamic VTs, coarticulation, and including a nasal cavity.

This study paves the way for a number of potential applications and further research on speech breathing. This may include conditions such as pathological speech and varying speech tasks, which are beyond the scope of this article. Along with its general importance for speech science, a better understanding of the acoustics of breath noises has implications for a variety of areas. In speech technology, it may help make automatic speech segmentation and alignment get better at differentiating breathing sounds from silent pauses and speech sounds and make breath noises in synthetic speech more natural (Braunschweiler & Chen, 2013). Breathing noise may be crucial to better understand breathing frequency in setups where only the audio signal is present (e.g., Romano et al., 2023), such as noninvasive or remote diagnosis. In clinical applications, respiratory frequency and noises could contribute to improving the automatic detection of pathologies, such as COVID-19, pathological cries and coughs in infancy, or vocal fold paralysis, or the emotional or cognitive state of the speaker. To test our acoustic-based findings, future studies should examine the supraglottal physiology of speech breathing. An approach similar to the studies on pause postures (Gick et al., 2004; Krivokapić et al., 2022) could be carried out. Future work may reveal if and to what degree what has been postulated as pause postures is influenced by inhalations. This would be particularly interesting for jaw and tongue movement: to study how far the jaw opens, whether the tongue actively produces an inhalation posture or only the jaw, to verify velar motion, and to see how much nasal contribution there is to inhalations (see Lester & Hoit, 2014). Ultimately, relating acoustic and kinematic aspects could help make inferences about a speaker's speech-breathing behavior from audio signals alone.

Data Availability Statement

The data sets generated and/or analyzed during this study are not publicly available, as consent for data publication was not collected via consent forms.

Author Contributions

Raphael Werner: Conceptualization (Equal), Data curation (Supporting), Formal analysis (Lead), Methodology

(Equal), Project administration (Equal), Software (Lead), Validation (Lead), Visualization (Lead), Writing – original draft (Lead), Writing – review & editing (Equal). **Susanne Fuchs:** Conceptualization (Equal), Data curation (Lead), Formal analysis (Supporting), Investigation (Equal), Methodology (Equal), Project administration (Equal), Software (Supporting), Supervision (Equal), Validation (Supporting), Visualization (Supporting), Writing –original draft (Supporting), Writing –review & editing (Equal). **Jürgen Trouvain:** Conceptualization (Equal), Data curation (Supporting), Funding acquisition (Lead), Methodology (Equal), Project administration (Equal), Resources (Lead), Supervision (Lead), Writing – review & editing (Equal). **Steffen Kürbis:** Conceptualization (Equal), Data curation (Lead), Formal analysis (Supporting), Investigation (Equal), Methodology (Equal), Project administration (Equal), Software (Supporting), Supervision (Supporting), Validation (Supporting), Writing – review & editing (Equal). **Bernd Möbius:** Conceptualization (Equal), Funding acquisition (Lead), Methodology (Equal), Project administration (Equal), Resources (Lea), Supervision (Lead), Writing – review & editing (Equal). **Peter Birkholz:** Conceptualization (Equal), Funding acquisition (Supporting), Investigation (Equal), Methodology (Equal), Project administration (Equal), Resources (Supporting), Supervision (Supporting), Writing – review & editing (Equal).

Acknowledgments

This research was funded in part by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project ID MO 597/10-1 (recipient: Bernd Möbius) and TR 468/3-1 (recipient: Jürgen Trouvain). We would like to thank Amélie Rochet-Capellan for giving us access to the female speaker data set and Michael Jessen and the Bundeskriminalamt, Department of Language and Audio (KT34), for providing us with the Pool 2010 corpus. We would like to thank Beeke Muhlack for her help with annotations.

References

Ahmed, N., Natarajan, T., & Rao, K. R. (1974). Discrete cosine transform. *IEEE Transactions on Computers, C-23*(1), 90–93. https://doi.org/10.1109/T-C.1974.223784

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Birkholz, P. (2013). Modeling consonant-vowel coarticulation for articulatory speech synthesis. *PLOS ONE, 8*(4), Article e60603. https://doi.org/10.1371/journal.pone.0060603

Birkholz, P., Gabriel, F., Kürbis, S., & Echternach, M. (2019). How the peak glottal area affects linear predictive coding-based formant estimates of vowels. *The Journal of the Acoustical Society of America, 146*(1), 223–232. https://doi.org/10.1121/1.5116137

Birkholz, P., Häsner, P., & Kürbis, S. (2022). Acoustic comparison of physical vocal tract models with hard and soft walls. *ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing*, 8242–8246. https://doi.org/10.1109/ICASSP43922.2022.9746611

Birkholz, P., Kürbis, S., Stone, S., Häsner, P., Blandin, R., & Fleischer, M. (2020). Printable 3D vocal tract shapes from MRI data and their acoustic and aerodynamic properties. *Scientific Data, 7*(1), 255–216. https://doi.org/10.1038/s41597-020-00597-w

Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer* [Computer software]. http://www.praat.org/

Braunschweiler, N., & Chen, L. (2013). Automatic detection of inhalation breath pauses for improved pause modelling in HMM-TTS. *Proceedings of the 8th ISCA Workshop on Speech Synthesis (SSW 8)*, 1–6.

Catford, J. C. (1977). *Fundamental problems in phonetics*. Edinburgh University Press.

Catford, J. C. (2001). *A practical introduction to phonetics* (2nd ed.). Oxford University Press.

Chen, Z., Li, M., Wang, R., Sun, W., Liu, J., Li, H., & Wang, X. (2022). Diagnosis of COVID-19 via acoustic analysis and artificial intelligence by monitoring breath sounds on smartphones. *Journal of Biomedical Informatics, 130,* Article 104078. https://doi.org/10.1016/j.jbi.2022.104078

Conrad, B., & Schönle, P. (1979). Speech and respiration. *Archiv für Psychiatrie und Nervenkrankheiten, 226*(4), 251–268. https://doi.org/10.1007/BF00342238

Conrad, B., Thalacker, S., & Schönle, P. (1983). Speech respiration as an indicator of integrative contextual processing. *Folia Phoniatrica et Logopaedica, 35*(5), 220–225. https://doi.org/10.1159/000265766

Dang, J., & Honda, K. (2002). Estimation of vocal tract shapes from speech sounds with a physiological articulatory model. *Journal of Phonetics, 30*(3), 511–532. https://doi.org/10.1006/jpho.2002.0167

Eklund, R. (2008). Pulmonic ingressive phonation: Diachronic and synchronic characteristics, distribution and function in animal and human sound production and in human speech. *Journal of the International Phonetic Association, 38*(3), 235–324. https://doi.org/10.1017/S0025100308003563

Fink, B. R. (1974). Folding mechanism of the human larynx. *Acta Oto-Laryngologica, 78*(1–6), 124–128. https://doi.org/10.3109/00016487409126336

Fuchs, S., Koenig, L., & Werner, R. (2023). Respiratory and supralaryngeal effects on speech breathing noise across loudness conditions and speaking tasks. *Proceedings of the 20th International Congress of Phonetic Sciences, Prague*, 1127–1131.

Fuchs, S., Petrone, C., Krivokapić, J., & Hoole, P. (2013). Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics, 41*(1), 29–47. https://doi.org/10.1016/j.wocn.2012.08.007

Fuchs, S., & Rochet-Capellan, A. (2021). The respiratory foundations of spoken language. *Annual Review of Linguistics, 7*(1), 13–30. https://doi.org/10.1146/annurev-linguistics-031720-103907

Gick, B., Wilson, I., & Derrick, D. (2013). *Articulatory phonetics*. Wiley.

Gick, B., Wilson, I., Koch, K., & Cook, C. (2004). Language-specific articulatory settings: Evidence from inter-utterance rest position. *Phonetica, 61*(4), 220–233. https://doi.org/10.1159/000084159

Goldman-Eisler, F. (1955). Speech-breathing activity—A measure of tension and affect during interviews. *British Journal of Psychology, 46*(1), 53–63. https://doi.org/10.1111/j.2044-8295.1955.tb00524.x

Hallani, M., Wheatley, J. R., & Amis, T. C. (2008). Enforced mouth breathing decreases lung function in mild asthmatics. *Respirology, 13*(4), 553–558. https://doi.org/10.1111/j.1440-1843.2008.01300.x

Hanna, N., Smith, J., & Wolfe, J. (2016). Frequencies, bandwidths and magnitudes of vocal tract and surrounding tissue resonances, measured through the lips during phonation. *The Journal of the Acoustical Society of America, 139*(5), 2924–2936. https://doi.org/10.1121/1.4948754

Hanna, N., Smith, J., & Wolfe, J. (2018). How the acoustic resonances of the subglottal tract affect the impedance spectrum measured through the lips. *The Journal of the Acoustical Society of America, 143*(5), 2639–2650. https://doi.org/10.1121/1.5033330

Harari, D., Redlich, M., Miri, S., Hamud, T., & Gross, M. (2010). The effect of mouth breathing versus nasal breathing on dentofacial and craniofacial development in orthodontic patients. *The Laryngoscope, 120*(10), 2089–2093. https://doi.org/10.1002/lary.20991

Heck, D. H., McAfee, S. S., Liu, Y., Babajani-Feremi, A., Rezaie, R., Freeman, W. J., Wheless, J. W., Papanicolaou, A. C., Ruszinkó, M., Sokolov, Y., & Kozma, R. (2017). Breathing as a fundamental rhythm of brain function. *Frontiers in Neural Circuits, 10,* Article 115. https://doi.org/10.3389/fncir.2016.00115

Heeren, W. F. L. (2015). Vocalic correlates of pitch in whispered versus normal speech. *The Journal of the Acoustical Society of America, 138*(6), 3800–3810. https://doi.org/10.1121/1.4937762

Hirschberg, J. (1980). Acoustic analysis of pathological cries, stridors and coughing sounds in infancy. *International Journal of Pediatric Otorhinolaryngology, 2*(4), 287–300. https://doi.org/10.1016/0165-5876(80)90034-8

Hixon, T. J., Weismer, G., & Hoit, J. D. (2020). *Preclinical speech science: Anatomy, physiology, acoustics, and perception*. Plural.

Jannedy, S., & Weirich, M. (2017). Spectral moments vs discrete cosine transformation coefficients: Evaluation of acoustic measures distinguishing two merging German fricatives. *The Journal of the Acoustical Society of America, 142*(1), 395–405. https://doi.org/10.1121/1.4991347

Jessen, M., Köster, O., & Gfroerer, S. (2005). Influence of vocal effort on average and variability of fundamental frequency. *International Journal of Speech, Language and the Law, 12*(2), 174–213. https://doi.org/10.1558/sll.2005.12.2.174

Kienast, M., & Glitza, F. (2003). Respiratory sounds as an idiosyncratic feature in speaker recognition. *International Congress of Phonetic Sciences*, 1607–1610.

Krivokapić, J., Styler, W., & Byrd, D. (2022). The role of speech planning in the articulation of pauses. *The Journal of the Acoustical Society of America, 151*(1), 402–413. https://doi.org/10.1121/10.0009279

Kuhlmann, L. L., & Iwarsson, J. (2021). Effects of speaking rate on breathing and voice behavior. *Journal of Voice*. Advance online publication. https://doi.org/10.1016/j.jvoice.2021.09.005

Lenth, R. V. (2021). *emmeans: Estimated marginal means, aka least-squares means (R package Version 1.6.1)* [Computer software].

Lester, R. A., & Hoit, J. D. (2014). Nasal and oral inspiration during natural speech breathing. *Journal of Speech, Language, and Hearing Research, 57*(3), 734–742. https://doi.org/10.1044/1092-4388(2013/13-0096)

Lulich, S. M. (2010). Subglottal resonances and distinctive features. *Journal of Phonetics, 38*(1), 20–32. https://doi.org/10.1016/j.wocn.2008.10.006

Mitchell, H. L., Hoit, J. D., & Watson, P. J. (1996). Cognitive-linguistic demands and speech breathing. *Journal of Speech, Language, and Hearing Research, 39*(1), 93–104. https://doi.org/10.1044/jshr.3901.93

Nakano, T., Ogata, J., Goto, M., & Hiraga, Y. (2008). Analysis and automatic detection of breath sounds in unaccompanied singing voice. *Proceedings of the 10th International Conference on Music Perception and Cognition*, 387–390.

Ng, A. K., Koh, T. S., Baey, E., Lee, T. H., Abeyratne, U. R., & Puvanendran, K. (2008). Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea? *Sleep Medicine, 9*(8), 894–898. https://doi.org/10.1016/j.sleep.2007.07.010

Orlikoff, R. F., Baken, R. J., & Kraus, D. H. (1997). Acoustic and physiologic characteristics of inspiratory phonation. *The Journal of the Acoustical Society of America, 102*(3), 1838–1845. https://doi.org/10.1121/1.420090

Patri, J., Diard, J., & Perrier, P. (2015). Optimal speech motor control and token-to-token variability: A Bayesian modeling approach. *Biological Cybernetics, 109*(6), 611–626. https://doi.org/10.1007/s00422-015-0664-4

Perl, O., Ravia, A., Rubinson, M., Eisen, A., Soroka, T., Mor, N., Secundo, L., & Sobel, N. (2019). Human non-olfactory cognition phase-locked with inhalation. *Nature Human Behaviour, 3*(5), 501–512. https://doi.org/10.1038/s41562-019-0556-z

Ramanarayanan, V., Bresch, E., Byrd, D., Goldstein, L., & Narayanan, S. S. (2009). Analysis of pausing behavior in spontaneous speech using real-time magnetic resonance imaging of articulation. *The Journal of the Acoustical Society of America, 126*(5), EL160–EL165. https://doi.org/10.1121/1.3213452

Rasskazova, O., Mooshammer, C., & Fuchs, S. (2019). Temporal coordination of articulatory and respiratory events prior to speech initiation. *Interspeech 2019*, 884–888. https://doi.org/10.21437/Interspeech.2019-2876

Rochet-Capellan, A., & Fuchs, S. (2013). Changes in breathing while listening to read speech: The effect of reader and speech mode. *Frontiers in Psychology, 4,* Article 906. https://doi.org/10.3389/fpsyg.2013.00906

Romano, C., Nicolò, A., Innocenti, L., Bravi, M., Miccinilli, S., Sterzi, S., Sacchetti, M., Schena, E., & Massaroni, C. (2023). Respiratory rate estimation during walking and running using breathing sounds recorded with a microphone. *Biosensors, 13*(6), Article 637. https://doi.org/10.3390/bios13060637

RStudio Team. (2022). *RStudio: Integrated development environment for R*. http://www.rstudio.com/

Ruinskiy, D., & Lavner, Y. (2007). An effective algorithm for automatic detection and exact demarcation of breath sounds in speech and song signals. *IEEE Transactions on Audio, Speech and Language Processing, 15*(3), 838–850. https://doi.org/10.1109/TASL.2006.889750

Salomoni, S., Van Den Hoorn, W., & Hodges, P. (2016). Breathing and singing: Objective characterization of breathing patterns in classical singers. *PLOS ONE, 11*(5), Article e0155084. https://doi.org/10.1371/journal.pone.0155084

Scheinherr, A., Bailly, L., Boiron, O., Lagier, A., Legou, T., Pichelin, M., Caillibotte, G., & Giovanni, A. (2015). Realistic glottal motion and airflow rate during human breathing. *Medical Engineering and Physics, 37*(9), 829–839. https://doi.org/10.1016/j.medengphy.2015.05.014

Scobbie, J. M., Schaeffler, S., & Mennen, I. (2011). Audible aspects of speech preparation. *International Congress of Phonetic Sciences (ICPhS)*, 1782–1785.

Székely, É., Henter, G. E., & Gustafson, J. (2019). Casting to corpus: Segmenting and selecting spontaneous dialogue for TTS with a CNN-LSTM speaker-dependent breath detector. *IEEE International Conference on Acoustics, Speech and*

Signal Processing (ICASSP), 6925–6929. https://doi.org/10.1109/ICASSP.2019.8683846

Ungeheuer, G. (1969). *Elemente einer akustischen Theorie der Vokalartikulation* [Elements of an acoustic theory of vowel articulation]. Springer.

Werner, R., Fuchs, S., Trouvain, J., & Möbius, B. (2021). Inhalations in speech: Acoustic and physiological characteristics. *Interspeech 2021*, 3186–3190. https://doi.org/10.21437/Interspeech.2021-1262

Werner, R., Trouvain, J., Fuchs, S., & Möbius, B. (2021). Exploring the presence and absence of inhalation noises when speaking and when listening. *12th International Seminar on Speech Production (ISSP)*, 214–217.

Winkelmann, R., Jaensch, K., Cassidy, S., & Harrington, J. (2021). *emuR: Main package of the EMU speech database management system (R package Version 2.3.0)* [Computer software].

Winkworth, A. L., Davis, P. J., Ellis, E., & Adams, R. D. (1994). Variability and consistency in speech breathing during reading: Lung volumes, speech intensity, and linguistic factors. *Journal of Speech and Hearing Research, 37*(3), 535–556. https://doi.org/10.1044/jshr.3703.535

Zöllner, A., Mooshammer, C., Rasskazova, O., & Fuchs, S. (2021). Breathing affects reaction time in simple and delayed naming tasks. *12th International Seminar on Speech Production (ISSP)*, 218–221.