


Article

Pareto-Optimized Non-Negative Matrix Factorization Approach to the Cleaning of Alaryngeal Speech Signals

Rytis Maskeliūnas ^{1,*}, Robertas Damaševičius ¹, Audrius Kulikajevas ¹, Kipras Pribuišis ², Nora Ulozaitė-Stanienė ² and Virgilijus Uloza ²

¹ Faculty of Informatics, Kaunas University of Technology, 44249 Kaunas, Lithuania; robertas.damasevicius@ktu.lt (R.D.); audrius.kulikajevas@ktu.lt (A.K.)

² Department of Otorhinolaryngology, Academy of Medicine, Lithuanian University of Health Sciences, 44240 Kaunas, Lithuania; kipras.pribuisis@ismuni.lt (K.P.); nora.ulozaite@ismuni.lt (N.U.-S.); virgilijus.ulozas@ismuni.lt (V.U.)

* Correspondence: rytis.maskeliunas@ktu.lt

Simple Summary: This paper introduces a new method for cleaning impaired speech by combining Pareto-optimized deep learning with Non-negative Matrix Factorization (NMF). The approach effectively reduces noise in impaired speech while preserving the desired speech quality. The method involves calculating the spectrogram of a noisy voice clip, determining a noise threshold, computing a noise-to-signal mask, and smoothing it to avoid abrupt transitions. Using a Pareto-optimized NMF, the modified spectrogram is decomposed into basis functions and weights, allowing for reconstruction of the clean speech spectrogram. The final result is a noise-reduced waveform achieved by inverting the clean speech spectrogram. Experimental results validate the method's effectiveness in cleaning alaryngeal speech signals, indicating its potential for real-world applications.

Abstract: The problem of cleaning impaired speech is crucial for various applications such as speech recognition, telecommunication, and assistive technologies. In this paper, we propose a novel approach that combines Pareto-optimized deep learning with non-negative matrix factorization (NMF) to effectively reduce noise in impaired speech signals while preserving the quality of the desired speech. Our method begins by calculating the spectrogram of a noisy voice clip and extracting frequency statistics. A threshold is then determined based on the desired noise sensitivity, and a noise-to-signal mask is computed. This mask is smoothed to avoid abrupt transitions in noise levels, and the modified spectrogram is obtained by applying the smoothed mask to the signal spectrogram. We then employ a Pareto-optimized NMF to decompose the modified spectrogram into basis functions and corresponding weights, which are used to reconstruct the clean speech spectrogram. The final noise-reduced waveform is obtained by inverting the clean speech spectrogram. Our proposed method achieves a balance between various objectives, such as noise suppression, speech quality preservation, and computational efficiency, by leveraging Pareto optimization in the deep learning model. The experimental results demonstrate the effectiveness of our approach in cleaning alaryngeal speech signals, making it a promising solution for various real-world applications.

Keywords: alaryngeal; voice quality; voice cleaning; voice disorders; Pareto optimization



Citation: Maskeliūnas, R.; Damaševičius, R.; Kulikajevas, A.; Pribuišis, K.; Ulozaitė-Stanienė, N.; Uloza, V. Pareto-Optimized Non-Negative Matrix Factorization Approach to the Cleaning of Alaryngeal Speech Signals. *Cancers* **2023**, *15*, 3644. <https://doi.org/10.3390/cancers15143644>

Academic Editors: Giulia Marvaso and Federico Mastroleo

Received: 14 June 2023

Revised: 9 July 2023

Accepted: 12 July 2023

Published: 16 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Laryngeal cancer remains the most common malignant tumor in the upper respiratory tract [1]. Despite the decreasing incidence, approximately 60% of patients present with stage III or IV disease at the initial workup [2,3]. Surgery or surgery combined with chemoradiotherapy remains the preferred treatment method for laryngeal cancer, offering an optimal 5-year survival rate [4,5]. The surgical treatment options are laryngeal-preserving or radical surgery. Laryngeal-preserving surgery options can range from endolaryngeal cordectomy

with a laser to partial removal of the larynx—partial laryngectomy. The complete removal of the larynx, also known as a total laryngectomy, is the radical option. The more advanced the disease, the more radical the treatment required to achieve remission. Advanced laryngeal cancer stages limit the treatment options that can be offered to patients. In most cases, the complete removal of the larynx—total laryngectomy—is advised. Such surgery leaves the patient without the larynx, the main part of the vocal apparatus, and their vocal function is significantly impaired. Long-term voice and speech function rehabilitation is required, often with unsatisfactory results. Total laryngectomy results in the complete and permanent separation of the upper and lower airways and requires the creation of a terminal tracheostoma to breathe. The complete removal of the larynx and lack of air movement through the mouth results in patients' total loss of phonatory function [6]. After the removal of the larynx, the patient has to rely on alaryngeal speech to communicate. Alaryngeal speech can be achieved in three ways: esophageal speech, an electrolarynx, or a tracheoesophageal prosthesis (TEP). Esophageal speech and an electrolarynx benefit from low maintenance and do not require additional surgery. The TEP outperforms both methods by providing better perceptual (voice quality and intelligibility) and acoustic (maximum phonation time, fundamental frequency, and intensity) speech outcomes [7]. A TEP can be implanted through a tracheoesophageal fistula formed during laryngectomy or later [8]. It functions as a one-way valve that allows the air to move from the trachea to the esophagus but keeps the food and liquids from entering the lungs (see Figure 1).

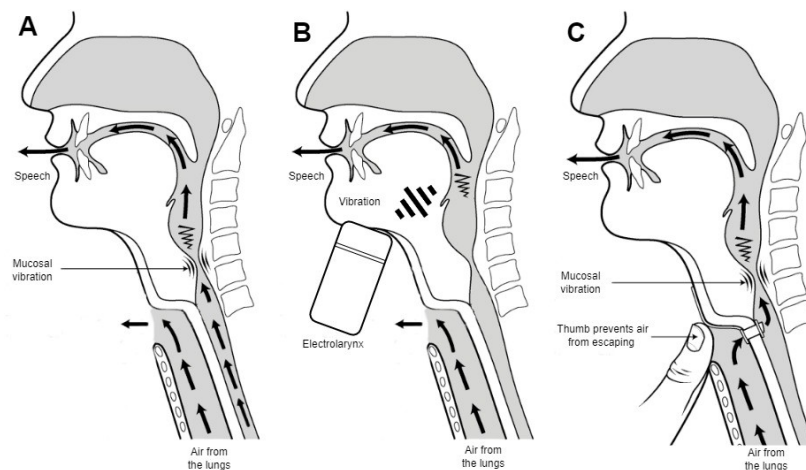


Figure 1. Types of speech production after total laryngectomy. (A) Esophageal speech: air is pulled in and released from the esophagus; (B) vibrations created by an electrolarynx; (C) the patient is occluding a tracheostoma to allow air to pass through the mouth. Adapted from Hurren 2015 [9].

The air moving through the TEP creates vibrations in the mucosa and generates speech [10]. The use of a pulmonary air supply to speak increases fluency and utterance lengths [11]. Despite its higher maintenance costs, the TEP is the preferred method for speech rehabilitation after total laryngectomy [12]. Alaryngeal (esophageal or TEP) speech is a patient's only verbal communication option after a total laryngectomy. Although the patient retains the ability to speak, the body begins to adapt and substitutes vocal folds with structures (aryepiglottic/ventricular folds, pharyngeal mucosa) that were not naturally intended for voice production. The downside of this adaptation is that the speech generated in this manner features frequent unintended phonatory breaks, frequency shifts, unvoiced segments, and high irregularity, and might be aperiodic (see Figure 2) [13]. It becomes even more problematic when the patient has to use the phone or speak in a loud environment, which may lead to social isolation [14,15]. The inability to communicate is most prominent in the early postoperative period before any speech rehabilitation occurs when patients have to rely on written text to communicate with their physician and family.

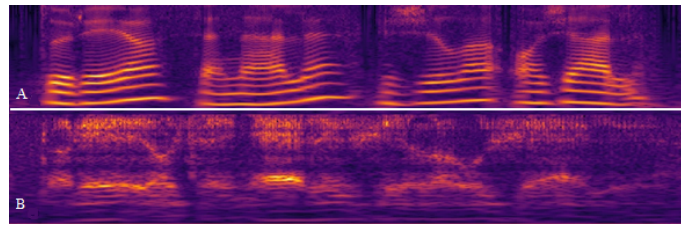


Figure 2. (A) Normal speech; (B) Speech after total laryngectomy with a tracheoesophageal prosthesis. Cochleagram deterioration and blurring between different words caused by aperiodicity, phonatory breaks, and additive noise.

It becomes even more problematic when the patient has to use the phone or speak in a loud environment, which may lead to social isolation [13,14]. The inability to communicate is most prominent in the early postoperative period before any speech rehabilitation occurs when patients have to rely on written text to communicate with their physician and family. Therefore, enhancing the signal quality of alaryngeal speech and improving a patient's speaking ability represent fundamental scientific/technical and clinical issues.

This may involve techniques such as breath control [16], pitch and tone modifications [17,18], and articulation exercises [19], including spectral subtraction [20], Wiener filtering [21], and statistical prediction model-based [22] or machine learning-based approaches [23]. However, these traditional methods often suffer from drawbacks such as introducing artifacts, suppressing the desired speech components, or being computationally expensive. With the advent of deep learning, several new approaches have been proposed that leverage the power of neural networks to address the limitations of traditional methods. Among these, non-negative matrix factorization (NMF) [24] has gained significant attention for its ability to represent non-negative data such as audio spectrograms as a linear combination of basis functions [25].

In this paper, we propose a novel approach for cleaning impaired speech signals by combining Pareto-optimized deep learning with NMF. Our method aims to balance various objectives, such as noise suppression, speech quality preservation, and computational efficiency, by leveraging Pareto optimization in a deep learning model. By incorporating Pareto optimization, we ensure that the trade-offs between different objectives are optimally balanced, ultimately improving the performance of the noise-reduction process.

The proposed method consists of several steps. First, we calculate the spectrogram of a noisy voice clip and extract its frequency statistics. Based on the desired noise sensitivity, a threshold is calculated to distinguish between the signal and noise components in the spectrogram. Next, we compute the noise-to-signal mask and smooth it to avoid abrupt transitions in noise levels. The modified spectrogram is then obtained by applying the smoothed mask to the signal spectrogram. To further enhance the speech signal, we employ a Pareto-optimized NMF to decompose the modified spectrogram into basis functions and corresponding weights. These basis functions and weights are learned to best represent the clean speech signal while achieving a balance between various objectives. Finally, the clean speech spectrogram is reconstructed using the learned basis functions and weights, and a noise-reduced waveform is obtained by inverting the clean speech spectrogram.

The main contributions of this paper are as follows:

- We propose a novel method for cleaning impaired speech signals by combining Pareto-optimized deep learning with NMF, addressing the limitations of traditional speech enhancement techniques.
- We introduce a smoothing technique for the noise-to-signal mask to avoid abrupt transitions in noise levels, resulting in a more natural-sounding output signal.
- We demonstrate the effectiveness of our approach through a series of experiments, showing significant improvements in speech quality and intelligibility compared to traditional methods.

The remainder of this paper is organized as follows. Section 2 provides a review of the related works in the field of speech enhancement. In Section 3, we describe our proposed method. Section 4 presents the experimental setup and results, followed by a discussion of the findings. Finally, Section 5 concludes the paper and suggests future research directions.

2. Review of State-of-the-Art Works

This overview of related works aims to help the reader explore the various approaches to improving speech intelligibility and quality for individuals with speech disorders. Various techniques, such as clear speech variants, adaptive filter structures, deep learning models, and speech enhancement algorithms, have been investigated to address the challenges in speech enhancement. These studies demonstrate the potential of different methods, including instruction-based interventions, signal processing, and machine learning techniques, to enhance speech intelligibility and quality across various disorders and conditions. The findings may help the reader better understand the complex relationship between speech impairments and the effectiveness of different approaches in overcoming these challenges, ultimately improving communication for affected individuals.

2.1. Assessing Speech-Signal Impairments

Evaluating the quality and intelligibility of alaryngeal speech can be difficult for several reasons [26]. First, evaluating speech quality is inherently subjective, as different people may have different opinions on what constitutes good or clear speech. Evaluators may also have biases or preconceived notions about alaryngeal speech, which can affect their judgment [27]. Second, alaryngeal speech can be complex and variable, depending on the individual's chosen method of alaryngeal speech and their level of proficiency [28], among other factors. Evaluators may need to consider multiple aspects of speech, such as pitch, tone, articulation, and prosody, which can make the evaluation more challenging [29]. Third, there is a limited amount of training data available for evaluating alaryngeal speech, as it is a relatively rare condition [30]. This can make it difficult to develop standardized evaluation methods and norms for different types of alaryngeal speech [31]. Finally, alaryngeal speech can vary widely between individuals, depending on factors such as age, gender, health status, and other individual characteristics [32]. This variability can make it difficult to develop standardized evaluation methods that are applicable to all individuals who have undergone a laryngectomy [33].

Numerous researchers have stressed the significance of selecting relevant characteristics for differentiating damaged speech [34–36]. Some researchers have investigated how speech difficulties caused by cerebral palsy and hearing impairment affect prosody, pronunciation, and voice quality. According to their findings, these factors are statistically significant for increasing the detection ability of impaired talks, with voice quality being the strongest discriminative feature for identifying speech intelligibility in damaged speech. Malini and Chandrakala [37] suggested a regularized self-representation-based compact supervector technique for assessing the intelligibility of damaged speech. On the UA-SPEECH database, their approach outperformed other methods such as hybrid GMM/SVM, supervector, x-vector, i-vector, and bag-of-models-based approaches. Albaqshi and Sagheer [38] emphasized the difficulties in dysarthric speech recognition owing to incomprehensible speech, irregular phoneme articulation, and data scarcity. Bessell et al. [39] found that a changed accent has a slower speech pace, greater consonant and vowel length, syllable-timed rhythm, and other characteristics. Moon et al. [40] sought to define the speech patterns of those suffering from hepatic encephalopathy as a possible diagnostic and monitoring tool. The subjects' maintained and damaged speech patterns, on the other hand, did not follow patterns normally linked with organic brain problems, suggesting that left-handed preference may contribute to distinctions between singing and reading vs. recitation, repetition, and spontaneous speaking. This is also often the case after an ischemic stroke. De Cock et al. [41] investigated speech features, dysarthria type, and severity, showing that unilateral upper motor neuron dysarthria is the most common type,

with the majority of subjects having mild dysarthria. Similarly, Rowe et al. [42] found that variable expressions of dysarthria may impact speech performance, whereas Stipanovic and Tjaden [43] found the least detectable change in sentence intelligibility in speakers with multiple sclerosis and Parkinson's disease. Rosdi et al. [44] presented fuzzy Petri nets to increase the classification accuracy of speech-intelligibility detection systems. Maskeliunas et al. suggested applying a convolutional network to help classify and assess impaired speech signals [45]. Kim et al. [46] used one- and two-dimensional convolutional neural networks to classify alaryngeal speech. Feng et al. [47] found that acoustic investigations can reveal that impaired speech has a substantially shorter voice start time for aspirated consonants, as well as a smaller vowel spacing. Vieira et al. [48] presented a non-intrusive voice-quality classifier based on the tree convolutional neural network for measuring user satisfaction with speech communication platforms. Poncelet et al. [49] suggested using an end-to-end spoken language understanding system that can be trained by the user through demonstrations and can translate impaired speech directly into semantics.

Numerous speech recognition-oriented techniques can also be used to help detect and assess speech impairment [50,51]. Gupta et al. [52] suggested a residual network-based approach for detecting dysarthria severity level based on short speech segments, whereas Latha et al. [53] employed deep learning and several acoustic cues to recognize dysarthric speech and generate discernible speech. Vishnika Veni and Chandrakala [54] researched the application of the deep neural network-hidden Markov model and lattice maximum mutual information technique for the successful identification of damaged speech. In [55], the authors suggested a histogram of states-based strategy for learning compact and discriminative embeddings for dysarthric voice detection using the deep neural network-hidden Markov model. Srinivasan et al. [56] proposed a multi-view representation-based disordered speech recognition system based on auditory image-based features and cepstral characteristics, showing improved performance in recognizing very low intelligibility words compared to conventional methods. Chandrakala et al. [57] presented a bag-of-models (BoM)-based approach that uses adjusted Gaussian mixture model (AGMM)-based embeddings for impaired speech-intelligibility evaluation. They tested the method on two datasets and discovered that it outperformed the supervector, hybrid GMM/SVM, i-vector, and x-vector-based techniques in terms of prediction error and reliability for intelligibility-level evaluation and score predictions. Fu et al. [58] created a Sch-net neural network built on a convolutional neural network for end-to-end schizophrenia speech identification using deep learning techniques, implying that it has the potential to help in the diagnosis of a particular language disability. Marini et al. [59] verified the efficacy of a speech analysis approach for dysarthria speakers by modifying the size and shift parameters of the spectral analysis window to increase ASR system performance.

2.2. Algorithms for Alaryngeal Speech Enhancement

The majority of voice restoration treatments result in hushed and monotonous speech. Aside from reduced intelligibility, this type of speech lacks expressiveness and naturalness due to (a) a lack of pitch, which results in whispered speech, and (b) artificial pitch production, which results in monotone speech. Algorithms for alaryngeal speech enhancement can be classified into two categories: classic digital signal processing (DSP) methods and methods based on artificial intelligence (AI) and machine learning (ML) [60].

The first category is the most popular as it includes filtering-based methods originally developed for noise reduction, as background noise can interfere with the clarity of alaryngeal speech [61]. DSP techniques, such as spectral subtraction, Wiener filtering, and adaptive filtering, can be used to reduce background noise and improve speech quality [62]. For example, Jaiswal et al. [63] suggested a concealed Wiener filter-based technique for voice augmentation to improve the common spectral subtraction algorithm. Pauline and Dhanalakshmi [64] presented an efficient adaptive filter structure for noise reduction in voice signals that utilized the least mean square (LMS) and normalized LMS algorithms. They evaluated the proposed filter model on both normal speech signals and speech signals

from Parkinson's disease patients. In terms of the SNR, MSE, and PSNR values, their filter model outperformed existing cascaded LMS filter models. Doi et al studied how the LPC spectrum of alaryngeal speech could be used to determine the impulse response of the vocal tract. Modified harmonic amplitudes calculated using the transformation function were interpolated at the desired harmonics of the target pitch, and the transformation function was then computed using the line spectral frequencies rather than the harmonic amplitudes [65]. Pauline et al. [66] presented cascaded adaptive filter construction for speech-signal de-noising, where the best variable-stage cascaded adaptive filter model outperformed existing cascaded filter architectures, with an output SNR that was 10–15 dB higher. Panda et al. [67] suggested using spectral subtraction to improve alaryngeal speech, which was modified by Hamed et al. to include the power of noise [68]. Wei suggested using the Mel Frequency Scale as an alternative [69]. Another approach is pitch and formant manipulation, as alaryngeal speech can have a monotonous or robotic quality due to a lack of natural pitch and formant variation. DSP techniques such as pitch shifting and formant manipulation can also be used to add more natural-sounding variation to speech [70]. Giri and Rayavarapu [71] presented a combined approach for modifying the key frequency, intensity, and speech rate of dysarthric speech by utilizing time-domain pitch-synchronous overlap. They discovered that the improvement in intelligibility was significant in speakers with low initial intelligibility and modest in speakers with high intelligibility. Additionally, there are methods for articulation enhancement, as alaryngeal speech can also suffer from poor articulation, making it difficult to distinguish between different sounds. This can be combated by utilizing dynamic range compression, and equalization can be used to enhance the clarity and intelligibility of specific consonant sounds [72]. Finally, prosody modification is common, as it can help process the patterns of stress, intonation, and rhythm in speech. Alaryngeal speech can sometimes lack the natural prosody found in normal speech and prosody modification can be used to add more natural-sounding patterns of stress, intonation, and rhythm to speech [73].

The second category includes AI and ML methods that can be used for alaryngeal speech enhancement. Currently, the most popular methods are deep learning models [74], such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs), which can be trained on large datasets of alaryngeal speech to learn patterns and relationships between speech features and speech quality. These models can be used to perform classic DSP tasks such as noise reduction, pitch and formant manipulation, and prosody modification [75]. Saleem et al. [76] suggested a computationally efficient deep learning model for improving noisy voice. For magnitude estimation, their model used a U-shaped fuzzy long short-term memory, which outperformed other deep learning models and significantly enhanced speech intelligibility and quality. In contrast to traditional GAN-based approaches [77], Santjago et al. [78] suggested using a speaker-dependent GAN to enhance generated speech. Others have proposed that an adversarial acoustic regression loss should be added to encourage better extraction of features at the discriminator and employ a two-step adversarial training schedule that serves as a warm-up and fine-tune sequence. Both the objective and subjective assessments indicated that these two enhancements improved speech reconstruction by better matching the original speaker's identity and naturalness [79]. Amarjuf combined the predicted phase with deep learning approaches to increase the overall quality [80]. Reinforcement learning can be applied to train speech enhancement systems that adapt to changing environments or input signals, as well as optimize speech enhancement systems based on a reward signal that reflects the quality of the enhanced speech [81]. Gaussian mixture models are also common, as they can work as a type of generative model that can be used to model the statistical distribution of speech features such as the spectral envelope or the fundamental frequency [82]. GMMs can also be used to separate speech from noise or modify the pitch and formant of speech. Ming et al. [83] suggested a hybrid technique that includes non-negative matrix factorization with GMM. According to Xui, the Gaussian mixture model approach is also beneficial

for detecting vocal nodules and laryngitis [84]. Support vector machines can be used to classify impaired speech signals into different categories such as normal speech or alaryngeal speech produced using different methods [85]. SVMs can also be used for noise reduction and speech enhancement [86]. Hidden Markov models can also be used as a generative model that models the statistical distribution of speech features over time, often to classify speech signals into different categories or generate new speech signals based on the statistical distribution of the input speech [87].

3. Materials and Methods

3.1. Dataset

Thirty native Lithuanian-speaking male patients surgically treated for histologically confirmed laryngeal cancer at the Lithuanian University of Health Sciences Department of Otorhinolaryngology provided speech samples for this study. The patients in this group had undergone a total laryngectomy with secondary TEP implantation [88,89]. These individuals were chosen because they had no larynx or vocal folds and relied solely on alaryngeal speech to communicate. The complete removal of the larynx and speech production using a TEP often result in distinct speech abnormalities with a fairly uniform functional speech handicap compared to neurodegenerative disorders, where speech patterns are more diverse and less distinct. The average age of the patients was 63.1 years (standard deviation = 28.8). The patients were free of common colds, upper respiratory infections, or other conditions that may have affected speech quality at the time of recording. Only male participants' speech samples were collected since advanced laryngeal cancer is less common in women, and recruiting an adequate number of female participants was not feasible. Endoscopic evaluation of the neopharynx, TEP canal, and trachea was performed prior to recording. Faulty or leaking prostheses were replaced prior to recording. This examination was carried out as part of standard clinical practice and contributed to the speech sample database exclusively containing speech samples from patients in remission. For at least six months following surgery, speech recordings were acquired. This ensured enough time for healing, speech adaptation, and rehabilitation [90].

Alaryngeal speech samples were recorded in a T-series quiet room (T-room, CA Tegner AB, Bromma, Sweden) using a D60S Dynamic Vocal microphone (AKG Acoustics, Vienna, Austria) placed 10.0 cm from the lips at a comfortable (about 90°) microphone-to-mouth angle. Two different speaking assignments were completed. The patient began by reading a phonetically balanced Lithuanian line: "Turėjo senelė žilą oželį" (Old grandma had a billy goat). The relative frequencies of the phonemes in the phrase were made as close as possible to the distribution of speech sounds in Lithuanian. The patient then counted from one to ten at a rate appropriate for their respiratory function. All speech activities were performed at a comfortable volume level and at the patient's own tempo. Speech was recorded at 44,100 samples per second and saved as uncompressed 16-bit waveform audio format files. Using Praat version 6.0.53, the recordings were manually prepared and contained no more than 300 ms of an unvoiced fragment at the beginning and conclusion of the recordings. To ensure the security of participants' personal data, serial numbers were assigned to the speech recordings.

3.2. Alaryngeal Speech Assessment

Several approaches were used to measure objective alaryngeal speech:

1. The artificial intelligence-based automated classifier for substitution voicing ResNet 118 was used to assign speech samples to the following classes: normal speech—Probability 0; speech with a single vocal fold—Probability 1; and alaryngeal speech with TEP—Probability 2 [91].
2. The acoustic parameter of alaryngeal speech (average voicing evidence (AVE), available in the AMPLEX software [92]) was utilized to compare the alaryngeal speech samples before and after optimization using Pareto-optimized NMF software. The AVE parameter describes the average voicing evidence and the degree of regular-

ity/periodicity in the voiced frames. Since the actual background frames are usually unvoiced, the analysis is performed on all frames, not just speech frames. This approach is more robust against possible errors of the speech/background classification, which is purely energy-based. In contrast, the voicing evidence is derived from analyzing all the sub-band signals created by the auditory model.

3. The AI-based acoustic substitution voicing index (ASVI) parameter [93] was employed to quantitatively evaluate the alaryngeal speech samples before and after optimization using Pareto-optimized NMF software. This parameter includes the constant combined with statistically significant parameters from ResNet 118 (Probability 0, Probability 1, and Probability 2) combined with the AVE and mean fundamental frequency. The possible ASVI values ranged from 0 to 30, with better speech quality indicated by higher scores.

3.3. Methodology

Our approach used Pareto-optimized deep learning to evaluate the possibility of cleaning the impaired speech. The approach started by calculating the spectrogram over the entire noisy voice clip, based on which the frequency statistics were calculated. Once the statistics were calculated, a threshold based on the desired noise sensitivity was then calculated. Afterward, a signal spectrogram was calculated based on the same input noisy voice clip, which, in combination with the calculated threshold, was then used to determine the noise-to-signal mask. The mask was then smoothed by applying a filter in both frequency and time to avoid sudden jumps in noise levels. Finally, the smoothed mask was then applied to the spectrogram of the signal and inverted creating a noise-reduced waveform.

3.3.1. Non-Negative Matrix Factorization (NMF)

Given a non-negative matrix $V \in \mathbb{R}_{\geq 0}^{m \times n}$, non-negative matrix factorization (NMF) aims to find two non-negative matrices $W \in \mathbb{R}_{\geq 0}^{m \times k}$ and $H \in \mathbb{R}_{\geq 0}^{k \times n}$ such that their product approximates the original matrix V :

$$V \approx WH \quad (1)$$

The objective is to minimize the distance between V and WH , typically measured by the Frobenius norm or another divergence measure:

$$\min_{W \geq 0, H \geq 0} \|V - WH\| \quad (2)$$

where $\|\cdot\|$ denotes the Frobenius norm or another divergence measure, and k is the desired dimensionality of the factorization (typically, $k \ll \min(m, n)$).

3.3.2. Pareto-Optimized Non-Negative Matrix Factorization (PONMF)

We define Pareto-optimized NMF as the problem of approximating a non-negative matrix V with the product of two non-negative matrices W and H , considering multiple objectives f_1, f_2, \dots, f_p . The Pareto-optimized NMF formulation seeks a solution that balances the trade-offs among these objectives, achieving a Pareto optimal solution where no objective can be improved without worsening at least one other objective.

Given a non-negative matrix $V \in \mathbb{R}_{\geq 0}^{m \times n}$, Pareto-optimized non-negative matrix factorization (NMF) aims to find two non-negative matrices $W \in \mathbb{R}_{\geq 0}^{m \times k}$ and $H \in \mathbb{R}_{\geq 0}^{k \times n}$ such that their product approximates the original matrix V :

$$V \approx WH \quad (3)$$

The objective is to find a Pareto optimal solution, considering multiple objectives f_1, f_2, \dots, f_p . A Pareto optimal solution is one where it is not possible to improve any

objective without worsening at least one other objective. The Pareto-optimized NMF can be formulated as:

$$\min_{W \geq 0, H \geq 0} (f_1(V, W, H), f_2(V, W, H), \dots, f_p(V, W, H)) \quad (4)$$

subject to Pareto optimality. Here, $f_i(V, W, H)$ represents the i -th objective such as minimizing the reconstruction error, promoting sparsity, or reducing computational complexity. The goal is to find a solution that balances the trade-offs among these objectives.

The first step is to calculate the spectrogram over the entire noisy voice clip to obtain a representation of the frequency spectrum of a signal over time. The noisy voice clip is windowed and its Fourier transform is calculated to obtain a spectrogram.

Once the spectrogram is calculated, frequency statistics are calculated to obtain a better understanding of the frequency distribution of the signal. This is achieved by calculating the mean and standard deviation of the magnitude of each frequency bin over time.

Based on the desired noise sensitivity, a threshold is calculated to distinguish between the signal and noise in the spectrogram. A signal spectrogram (see an example in Figure 3) is then calculated based on the same input noisy voice clip. This is achieved by windowing the noisy voice clip and taking its Fourier transform over time. The threshold calculated earlier is used to determine the noise-to-signal mask. The mask is a binary value for each frequency bin and time frame of the spectrogram, where 1 indicates the signal and 0 indicates noise. To avoid sudden jumps in noise levels, the mask is smoothed by applying a filter in both the frequency and time domains, making the noise-to-signal mask more continuous and less abrupt. Next, the smoothed mask is applied to the spectrogram of the signal, and the signal is inverted to create a noise-reduced waveform. This is achieved by multiplying the spectrogram of the signal with the smoothed mask and then taking the inverse Fourier transform over time to obtain the noise-reduced waveform. Then, a Pareto-optimized non-negative matrix factorization (NMF)-based method is applied to decompose the spectrogram into a set of basis functions and their corresponding weights. NMF-based methods for speech enhancement involve learning the basis functions and Pareto-optimized weights that best represent the clean speech signal and then using these to reconstruct the clean speech from a noisy input signal (Algorithm 1).

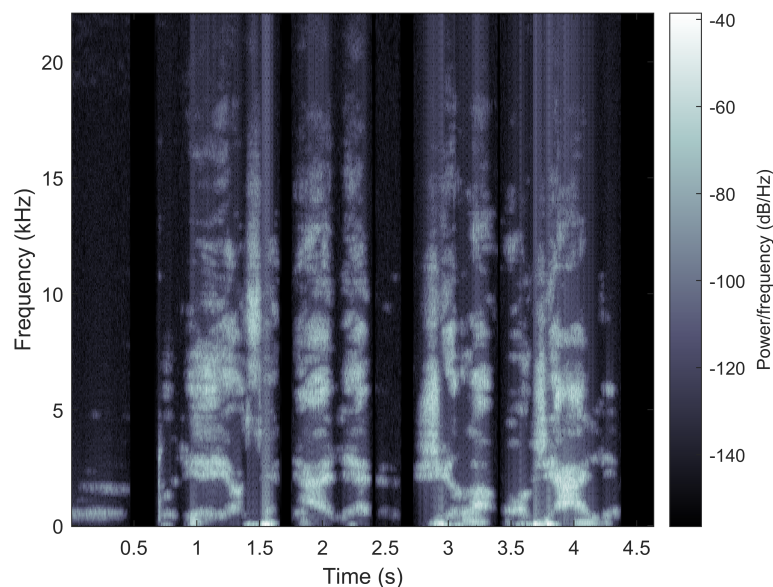


Figure 3. An example of a voice spectrogram.

Algorithm 1 Pareto-Optimized Deep Learning for Impaired Speech Cleaning**Require:** Noisy voice clip V **Ensure:** Noise-reduced waveform W

- 1: Calculate spectrogram S of noisy voice clip V
- 2: Compute frequency statistics F from spectrogram S
- 3: Calculate threshold T based on the desired noise sensitivity using frequency statistics F
- 4: Determine signal spectrogram S_{signal} using noisy voice clip V
- 5: Compute noise-to-signal mask M using threshold T and signal spectrogram S_{signal}
- 6: Smooth mask M by applying a filter in both the frequency and time domains to obtain smoothed mask M_{smooth}
- 7: Apply smoothed mask M_{smooth} to the spectrogram of signal S_{signal} to obtain modified spectrogram S_{mod}
- 8: Invert modified spectrogram S_{mod} to create noise-reduced waveform W
- 9: **return** W

3.3.3. Speech-Signal Cleaning

The updated approach to cleaning impaired speech using Pareto-optimized deep learning and non-negative matrix factorization (NMF) involves the following steps:

1. Calculate the spectrogram of the entire noisy voice clip. This is achieved by windowing the noisy voice clip and taking its Fourier transform over time to obtain a spectrogram, which is a representation of the frequency spectrum of a signal over time.
2. Compute the frequency statistics from the spectrogram. This is achieved by calculating the mean and standard deviation of the magnitude of each frequency bin over time. These statistics help in understanding the distribution and characteristics of the noise present in the voice clip.
3. Calculate a threshold based on the desired noise sensitivity. This threshold helps differentiate between the noise and signal components in the spectrogram.
4. Determine the signal spectrogram using the same input noisy voice clip. This is achieved by windowing the noisy voice clip and taking its Fourier transform over time.
5. Compute the noise-to-signal mask using the calculated threshold. The mask is a binary value for each frequency bin and time frame of the spectrogram, where 1 indicates the signal and 0 indicates noise.
6. Smooth the noise-to-signal mask by applying a filter in both the frequency and time domains. This helps avoid sudden jumps in noise levels and produces a more continuous and less abrupt mask.
7. Apply the smoothed mask to the spectrogram of the signal. This step effectively suppresses the noise components in the spectrogram while retaining the desired signal.
8. Decompose the modified spectrogram using Pareto-optimized non-negative matrix factorization (NMF). NMF-based methods for speech enhancement involve learning the basis functions and Pareto-optimized weights that best represent the clean speech signal.
9. Reconstruct the clean speech from the noisy input signal using the learned basis functions and Pareto-optimized weights.
10. Invert the reconstructed spectrogram to create a noise-reduced waveform. This final output is a cleaned version of the original impaired speech, with the noise components significantly reduced or removed.

3.3.4. Pareto-Optimized Deep Learning with NMF for Impaired Speech Cleaning

Using Pareto optimization in the deep learning model and incorporating NMF-based methods can ensure that the trade-offs between different objectives (e.g., noise suppression, speech quality, and computational efficiency) are balanced in the best possible way, ulti-

mately improving the performance of the noise-reduction process (Algorithm 2).

Algorithm 2 Pareto-Optimized Deep Learning with NMF for Impaired Speech Cleaning

Require: Noisy voice clip V

Ensure: Noise-reduced waveform W

- 1: Calculate spectrogram S of noisy voice clip V
 - 2: Compute frequency statistics F from spectrogram S
 - 3: Calculate threshold T based on the desired noise sensitivity using frequency statistics F
 - 4: Determine signal spectrogram S_{signal} using noisy voice clip V
 - 5: Compute noise-to-signal mask M using threshold T and signal spectrogram S_{signal}
 - 6: Smooth mask M by applying a filter in both the frequency and time domains to obtain smoothed mask M_{smooth}
 - 7: Apply smoothed mask M_{smooth} to the spectrogram of signal S_{signal} to obtain modified spectrogram S_{mod}
 - 8: Decompose modified spectrogram S_{mod} using Pareto-optimized non-negative matrix factorization (NMF) to obtain basis functions B and optimized weights W_{opt}
 - 9: Reconstruct clean speech spectrogram S_{clean} using basis functions B and optimized weights W_{opt}
 - 10: Invert clean speech spectrogram S_{clean} to create noise-reduced waveform W
-

4. Results

A statistically significant improvement in alaryngeal speech quality was observed because after applying Pareto-optimized NMF, the alaryngeal speech samples were reclassified into the lower speech disability category (see Table 1 and Figure 4).

Table 1. Evaluation results of original and optimized speech samples. Prob0—probability of healthy speech; Prob1—probability of speech with a single vocal fold; Prob2—probability of tracheoesophageal speech; AVE—average voicing evidence; ASVI—acoustic substitution voicing index; NMF—non-negative matrix factorization.

	Group	N	Mean	Std. Deviation	p
Probability 0	Original	75	4.09	19.51	0.001
	Pareto-optimized NMF	75	13.51	33.3	0.001
Probability 1	Original	75	56.18	48.66	0.454
	Pareto-optimized NMF	75	57.28	47.83	0.454
Probability 2	Original	75	39.73	47.9	0.08
	Pareto-optimized NMF	75	29.21	43.89	0.08
AVE	Original	75	0.81	0.11	0.34
	Pareto-optimized NMF	75	0.8	0.1	0.34
ASVI	Original	75	8.8	4.94	0.166
	Pareto-optimized NMF	75	10.17	6.09	0.166

To further test for improvements in the optimized speech samples, the Chi-squared test [94] was utilized to test if the proportion of speech recordings considered improved was large enough to be statistically significant. Only 4 out of 75 original speech recordings were classified as healthy, whereas 10 out of 75 were classified as healthy speech after optimization. This resulted in a statistically significant difference between the proportions ($p = 0.043$). These findings can be observed in Table 2.

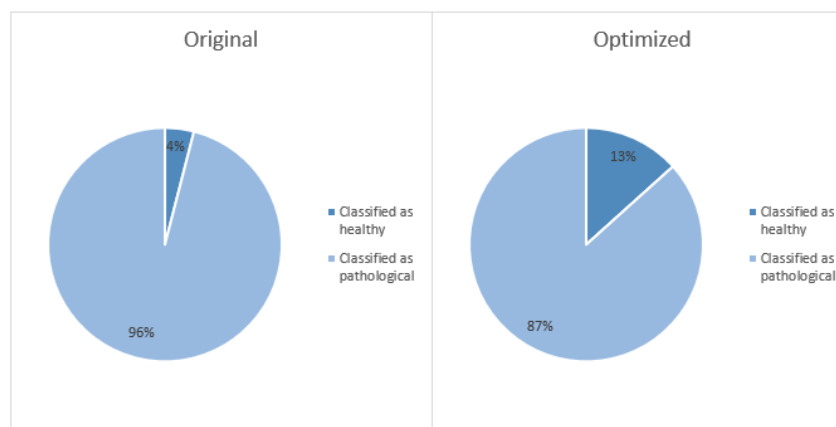


Figure 4. A pie chart illustrating the proportion of speech recordings classified as normal/pathological before and after speech-signal optimization.

Table 2. Comparison of original and optimized speech recordings. NMF—non-negative matrix factorization.

Group	Method	N	<i>p</i>	χ^2
Healthy speech	Original	4	4.0	0.043
	Pareto-optimized NMF	10	13.33	
Speech after laryngeal oncosurgery	Original	72	96.0	4.097
	Pareto-optimized NMF	65	86.67	

An example of the result of the alaryngeal speech-signal optimization is presented in Figure 5.

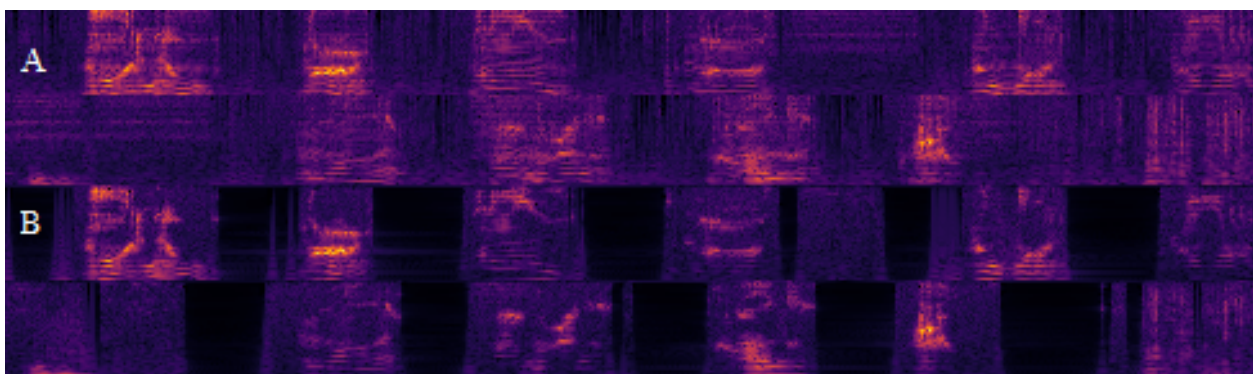


Figure 5. Cochleograms depicting a patient counting to ten with a tracheoesophageal prosthesis before (A) and after (B) speech-signal optimization. Less additive noise between separate numbers can be observed in the optimized cochleogram.

Table 3 presents the results of statistical tests (Levene’s test and t-test) performed on several groups of data (Probability 0, Probability 1, Probability 2, AVE, and ASVI).

Levene’s test for equality of variances checks whether the variances are equal across the groups. The null hypothesis is that the variances are equal. If the significance (sig.) is less than the threshold level (commonly 0.05), the null hypothesis is rejected, indicating that the variances are not equal. The choice between “equal variances assumed” and “equal variances not assumed” is determined by the results of Levene’s test. If the variances are found to be equal (sig. > 0.05 in Levene’s test), then we should refer to the t-test row for “equal variances assumed”. If the variances are not equal (sig. < 0.05 in Levene’s test), we should refer to the row “equal variances not assumed”. The significance for each group of data was as follows:

- Probability 0: sig. = 0.000, indicating that the variances were not equal across groups.
- Probability 1: sig. = 0.454, indicating that the variances were equal.
- Probability 2: sig. = 0.008, indicating that the variances were not equal.
- AVE: sig. = 0.340, indicating equal variances across groups.
- ASVI: sig. = 0.166, indicating equal variances across groups.

The *t*-test for equality of means checks whether the means of two groups are statistically significantly different. The null hypothesis is that the means are equal. If the significance (two-tailed) is less than the threshold level (commonly 0.05), the null hypothesis is rejected. The significance value for each group of data was as follows:

- Probability 0: sig. = 0.036 (for equal variances assumed) and 0.037 (for equal variances not assumed), indicating that the means of the two groups were significantly different.
- Probability 1: sig. = 0.890 (both cases), indicating that the means were not significantly different.
- Probability 2: sig. = 0.163 (both cases), indicating that the means are not significantly different.
- AVE: sig. = 0.750 (both cases), indicating that the means are not significantly different.
- ASVI: sig. = 0.133 (for equal variances assumed) and 0.134 (for equal variances not assumed), indicating that the means are not significantly different.

The mean difference, standard error difference, and 95% confidence interval of the difference provide further details on how the means of the two groups differed and the uncertainty surrounding that difference.

Table 3. Results of statistical tests.

		Levene's Test		t-Test for Equality of Means						
		F	Sig.	t	df	Sig. (2-Tailed)	Mean Difference	Std. Error Difference	95% Conf. Int.	
									Lower	Upper
Probability 0	Equal variances assumed	18.313	0.000	-2.113	148	0.036	-9.41893	4.45670	-18.22592	-0.61195
	Equal variances not assumed			-2.113	119.448	0.037	-9.41893	4.45670	-18.24330	-0.59457
Probability 1	Equal variances assumed	0.563	0.454	-0.139	148	0.890	-1.09627	7.87862	-16.66538	14.47284
	Equal variances not assumed			-0.139	147.956	0.890	-1.09627	7.87862	-16.66542	14.47288
Probability 2	Equal variances assumed	7.317	0.008	1.402	148	0.163	10.51547	7.50161	-4.30864	25.33957
	Equal variances not assumed			1.402	146.885	0.163	10.51547	7.50161	-4.30957	25.34050
AVE	Equal variances assumed	0.918	0.340	0.319	148	0.750	0.005560	0.017451	-0.028926	0.040046
	Equal variances not assumed			0.319	147.237	0.750	0.005560	0.017451	-0.028927	0.040047
ASVI	Equal variances assumed	1.941	0.166	-1.509	148	0.133	-1.36607	0.90525	-3.15495	0.42281
	Equal variances not assumed			-1.509	141.961	0.134	-1.36607	0.90525	-3.15558	0.42343

To summarize, as shown in Table 1, the mean AVE of the alaryngeal speech samples decreased from 81 to 80% after optimization. The AVE proportion remained statistically significant and unchanged in the samples before and after Pareto-NMF optimization. This is understandable and expected because the Pareto-optimized NMF approach removed background noise without artificially improving the quality of the alaryngeal speech recordings by filling in the unvoiced speech segments (pauses, intended phonatory breaks, etc.).

Lastly, the speech samples were evaluated using the ASVI, which represents the scale of the objective improvement of the alaryngeal speech signals when comparing the original and Pareto-optimized NMF alaryngeal speech recordings. Although the ASVI was higher in the group after optimization, the difference was not statistically significant. A description of the aforementioned evaluation can be found in Table 1.

5. Discussion

Speech is the complex result of several systems in the body working together. First, the respiratory tract must move air through the larynx and mouth. The vocal folds need to function correctly to produce voice. Speech is produced only when the articulation occurs in the pharynx and mouth and is then processed by the speaker's neural feedback loop, which helps correct the pitch and loudness. Finally, speech is used to communicate, so it has to be pleasant or, at the very least, intelligible to the listener [95]. Disturbances in any of these steps cause various levels of speech impairment.

Total laryngectomy patients often undergo speech rehabilitation programs to learn alternative methods of speech production such as esophageal speech, an electrolarynx, or tracheoesophageal speech with a voice prosthesis. These techniques can generate additional noise during speech production, thereby affecting speech quality and intelligibility. Implementing noise-reduction strategies can help mitigate this issue by improving the overall clarity and naturalness of the patient's speech [7]. However, a speech handicap becomes more problematic when the patient has to use the phone or speak in a loud environment, which may lead to social isolation [14,15].

The suggested Pareto–NMF optimization approach helps mitigate the additive and background noise problem that is common in alaryngeal speakers. The Pareto–NMF optimization removes additive and background noise without impacting the AVE. Although minuscule for a regular speaker, this improvement benefits the TEP speaker significantly. Firstly, total laryngectomy patients often face challenges in making their speech intelligible, especially in noisy environments. Excessive background noise can mask their already limited vocal output, making it difficult for listeners to understand them. By reducing the additional noise present in the environment, speech clarity and intelligibility can be improved, allowing patients to communicate more effectively. Secondly, speaking on the phone can be particularly challenging for individuals after laryngectomy [96]. Background noise, distortions, and limited vocal output can make it difficult for the listener to comprehend the speech. Unwanted noise can be minimized by implementing noise-reduction techniques, enabling clearer and more understandable phone conversations for laryngectomy patients.

Minimal Pareto–NMF optimization impact on speech benefits perfect TEP speakers more, as the spoken segments are largely unaltered and rely solely on the speaker's ability to speak clearly. Patients who have trouble articulating with a TEP could potentially benefit more from Pareto–NMF optimization combined with a speech enhancement model that addresses unvoiced segments, aperiodicity, and phonatory breaks that are more frequent in less experienced alaryngeal speakers.

A typical laryngeal cancer patient eligible for total laryngectomy and TEP rehabilitation is between 50 and 70 years of age and rarely has significant comorbidities [97,98]. After successful treatment, it is reasonable to expect at least a 40% 5-year survival rate. The combination of these conditions leads to a rather specific problem—a large group of patients that are functionally able to return to completely normal life or even the workforce but are held back by their speech disability. Alaryngeal speech enhancement techniques can help mitigate this problem and to allow complete rehabilitation and reintegration for patients after total laryngectomy.

6. Conclusions

Speech after surgical treatment for laryngeal cancer tends to suffer from aperiodicity, phonatory breaks, and additive noise [90]. These findings become more common as more laryngeal structures are removed. However, the adaptive capabilities of patients can result in vastly different acoustical outcomes despite undergoing identical surgery. This is reflected in the relatively high standard deviation observed when evaluating the ASVI of original and optimized speech samples. With this in mind, studies on acoustic speech after laryngeal oncosurgery should be carried out with a greater number of recordings.

Author Contributions: Conceptualization, V.U. and R.M.; Data curation, K.P.; Formal analysis, R.M., R.D., A.K., K.P., N.U.-S. and V.U.; Funding acquisition, V.U.; Investigation, K.P. and N.U.-S.; Methodology, K.P. and V.U.; Project administration, R.M. and V.U.; Resources, R.D., K.P. and V.U.; Software, A.K.; Supervision, R.M.; Validation, R.M., R.D. and K.P.; Visualization, R.D. and K.P.; Writing—original draft, R.M.; Writing—review and editing, R.M., K.P. and R.D. All authors have read and agreed to the published version of the manuscript.

Funding: This project received funding from the European Regional Development Fund (project no. 13.1.1-LMT-K-718-05-0027) under a grant agreement from the Research Council of Lithuania (LMTLT). This project was funded as a measure by the European Union in response to the COVID-19 pandemic.

Institutional Review Board Statement: This study was approved by the Kaunas Regional Ethics Committee for Biomedical Research (2022-04-20 No. BE-2-49).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available in this article.

Acknowledgments: The authors acknowledge the use of artificial intelligence tools for grammar checking and language improvement.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

NMF	non-negative matrix factorization
UA-SPEECH	sound dataset
GMM	Gaussian mixture model
SVM	support vector machine
AGMM	adjusted Gaussian mixture model
DSP	digital signal processing
AI	Artificial Intelligence
LMS	least mean square
SNR	signal-to-noise ratio
MSE	mean square error
PSNR	peak signal-to-noise ratio
LPC	linear predictive coding
CNN	convolutional neural network
RNN	recurrent neural network
GAN	generative adversarial network
TEP	tracheoesophageal prosthesis
Prob0	probability of healthy speech
Prob1	probability of speech with a single vocal fold
Prob2	probability of tracheoesophageal speech
AVE	average voicing evidence
ASVI	acoustic substitution voicing index

References

1. Steuer, C.E.; El-Deiry, M.; Parks, J.R.; Higgins, K.A.; Saba, N.F. An update on larynx cancer. *CA Cancer J. Clin.* **2016**, *67*, 31–50. [[CrossRef](#)] [[PubMed](#)]
2. Groome, P.A.; O’Sullivan, B.; Irish, J.C.; Rothwell, D.M.; Schulze, K.; Warde, P.R.; Schneider, K.M.; Mackenzie, R.G.; Hodson, D.I.; Hammond, J.A.; et al. Management and Outcome Differences in Supraglottic Cancer Between Ontario, Canada, and the Surveillance, Epidemiology, and End Results Areas of the United States. *J. Clin. Oncol.* **2003**, *21*, 496–505. [[CrossRef](#)] [[PubMed](#)]
3. Groome, P.; Schulze, K.; Keller, S.; Mackillop, W.; O’Sullivan, B.; Irish, J.; Bissett, R.; Dixon, P.; Eapen, L.; Gulavita, S.; et al. Explaining Socioeconomic Status Effects in Laryngeal Cancer. *Clin. Oncol.* **2006**, *18*, 283–292. [[CrossRef](#)]
4. Hoffman, H.T.; Porter, K.; Karnell, L.H.; Cooper, J.S.; Weber, R.S.; Langer, C.J.; Ang, K.K.; Gay, G.; Stewart, A.; Robinson, R.A. Laryngeal Cancer in the United States: Changes in Demographics, Patterns of Care, and Survival. *Laryngoscope* **2006**, *116*, 1–13. [[CrossRef](#)] [[PubMed](#)]

5. Caudell, J.J.; Gillison, M.L.; Maghami, E.; Spencer, S.; Pfister, D.G.; Adkins, D.; Birkeland, A.C.; Brizel, D.M.; Busse, P.M.; Cmelak, A.J.; et al. NCCN Guidelines® Insights: Head and Neck Cancers, Version 1.2022. *J. Natl. Compr. Cancer Netw.* **2022**, *20*, 224–234. [[CrossRef](#)]
6. Allegra, E.; Mantia, I.L.; Bianco, M.R.; Drago, G.D.; Fosse, M.C.L.; Azzolina, A.; Grillo, C.; Saita, V. Verbal performance of total laryngectomized patients rehabilitated with esophageal speech and tracheoesophageal speech: Impacts on patient quality of life. *Psychol. Res. Behav. Manag.* **2019**, *12*, 675–681. [[CrossRef](#)]
7. van Sluis, K.E.; van der Molen, L.; van Son, R.J.J.H.; Hilgers, F.J.M.; Bhairosing, P.A.; van den Brekel, M.W.M. Objective and subjective voice outcomes after total laryngectomy: A systematic review. *Eur. Arch. Oto-Rhino-Laryngol.* **2017**, *275*, 11–26. [[CrossRef](#)]
8. Chakravarty, P.D.; McMurrin, A.E.L.; Banigo, A.; Shakeel, M.; Ah-See, K.W. Primary versus secondary tracheoesophageal puncture: Systematic review and meta-analysis. *J. Laryngol. Otol.* **2017**, *132*, 14–21. [[CrossRef](#)]
9. Hurren, A.; Miller, N. Voice outcomes post total laryngectomy. *Curr. Opin. Otolaryngol. Head Neck Surg.* **2017**, *25*, 205–210. [[CrossRef](#)]
10. Kotby, M.; Hegazi, M.; Kamal, I.; el Dien, N.G.; Nassar, J. Aerodynamics of the Pseudo-Glottis. *Folia Phoniatr. Logop.* **2009**, *61*, 24–28. [[CrossRef](#)]
11. Brook, I.; Goodman, J.F. Tracheoesophageal Voice Prosthesis Use and Maintenance in Laryngectomees. *Int. Arch. Otorhinolaryngol.* **2020**, *24*, e535–e538. [[CrossRef](#)] [[PubMed](#)]
12. de Coul, B.M.R.O.; Hilgers, F.J.M.; Balm, A.J.M.; Tan, I.B.; van den Hoogen, F.J.A.; van Tinteren, H. A Decade of Postlaryngectomy Vocal Rehabilitation in 318 Patients. *Arch. Otolaryngol. Head Neck Surg.* **2000**, *126*, 1320. [[CrossRef](#)] [[PubMed](#)]
13. Dejonckere, P.H.; Bradley, P.; Clemente, P.; Cornut, G.; Crevier-Buchman, L.; Friedrich, G.; Heyning, P.V.D.; Remacle, M.; Woisard, V. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *Eur. Arch. Oto-Rhino-Laryngol.* **2001**, *258*, 77–82. [[CrossRef](#)] [[PubMed](#)]
14. Semple, C.; Parahoo, K.; Norman, A.; McCaughan, E.; Humphris, G.; Mills, M. Psychosocial interventions for patients with head and neck cancer. *Cochrane Database Syst. Rev.* **2013**. [[CrossRef](#)] [[PubMed](#)]
15. Uscher-Pines, L.; Sousa, J.; Raja, P.; Mehrotra, A.; Barnett, M.L.; Huskamp, H.A. Suddenly Becoming a “Virtual Doctor”: Experiences of Psychiatrists Transitioning to Telemedicine During the COVID-19 Pandemic. *Psychiatr. Serv.* **2020**, *71*, 1143–1150. [[CrossRef](#)]
16. Bohnenkamp, T.A. Postlaryngectomy Respiratory System and Speech Breathing. In *Clinical Care and Rehabilitation in Head and Neck Cancer*; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 103–117. [[CrossRef](#)]
17. Qian, Z.; Niu, H.; Wang, L.; Kobayashi, K.; Zhang, S.; Toda, T. Mandarin Electro-Laryngeal Speech Enhancement based on Statistical Voice Conversion and Manual Tone Control. In Proceedings of the 2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Tokyo, Japan, 14–17 December 2021; pp. 546–552.
18. Dinh, T.; Kain, A.; Samlan, R.; Cao, B.; Wang, J. Increasing the Intelligibility and Naturalness of Alaryngeal Speech Using Voice Conversion and Synthetic Fundamental Frequency. *Proc. Interspeech* **2020**, *2020*, 4781–4785. [[CrossRef](#)]
19. Graham, M.S. Strategies for Excelling With Alaryngeal Speech Methods. *Perspect. Voice Voice Disord.* **2006**, *16*, 25–32. [[CrossRef](#)]
20. Kabir, R.; Greenblatt, A.; Panetta, K.; Agaian, S. Enhancement of alaryngeal speech utilizing spectral subtraction and minimum statistics. In Proceedings of the 2008 International Conference on Machine Learning and Cybernetics, Kunming, China, 12–15 July 2008; Volume 7, pp. 3704–3709. [[CrossRef](#)]
21. Garg, A.; Sahu, O.P. Enhancement of speech signal using diminished empirical mean curve decomposition-based adaptive Wiener filtering. *Pattern Anal. Appl.* **2019**, *23*, 179–198. [[CrossRef](#)]
22. Wang, Q.; Du, X.; Gu, W. A Source-Filter Model-Based Unvoiced Speech Detector for Speech Coding. In Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013), Hangzhou, China, 22–23 March 2013; Atlantis Press: Dordrecht, The Netherlands, 2013. [[CrossRef](#)]
23. Huq, M.; Maskeliunas, R. Speech Enhancement Using Generative Adversarial Network (GAN). In *Hybrid Intelligent Systems*; Springer International Publishing: Berlin/Heidelberg, Germany, 2022; pp. 273–282. [[CrossRef](#)]
24. Sack, A.; Jiang, W.; Perlmutter, M.; Salanevich, P.; Needell, D. On Audio Enhancement via Online Non-Negative Matrix Factorization. In Proceedings of the 2022 56th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA, 9–11 March 2022; pp. 287–291. [[CrossRef](#)]
25. Wang, D.; Cui, J.; Wang, J.; Tan, H.; Xu, M. Convex Hull Convolutional Non-negative Matrix Factorization Based Speech Enhancement For Multimedia Communication. In Proceedings of the 2022 6th International Conference on Cryptography, Security and Privacy (CSP), Tianjin, China, 14–16 January 2022; pp. 138–142. [[CrossRef](#)]
26. Knollhoff, S.M.; Borrie, S.A.; Barrett, T.S.; Searl, J.P. Listener impressions of alaryngeal communication modalities. *Int. J. Speech-Lang. Pathol.* **2021**, *23*, 540–547. [[CrossRef](#)]
27. Mouret, F.; Crevier-Buchman, L.; Pillot-Loiseau, C. Intelligibility of pseudo-whispered speech after total laryngectomy. *Clin. Linguist. Phon.* **2022**, 1–17. [[CrossRef](#)]
28. Hui, T.F.; Cox, S.R.; Huang, T.; Chen, W.R.; Ng, M.L. The Effect of Clear Speech on Cantonese Alaryngeal Speakers’ Intelligibility. *Folia Phoniatr. Logop.* **2021**, *74*, 103–111. [[CrossRef](#)] [[PubMed](#)]
29. Aueworakhunanan, T.; Dechongkit, S.; Jeeraumporn, J.; Punkla, W. An Evaluation Pertaining to Esophageal Speech Outcomes in Alaryngeal Patients. *Ramathibodi Med. J.* **2022**, *45*, 16–24. [[CrossRef](#)]

30. Cao, B.; Teplansky, K.; Sebkhii, N.; Bhavsar, A.; Inan, O.; Samlan, R.; Mau, T.; Wang, J. Data Augmentation for End-to-end Silent Speech Recognition for Laryngectomees. *Proc. Interspeech* **2022**, *2022*, 3653–3657. [[CrossRef](#)]
31. Kent, R.D.; Kim, Y.; mei Chen, L. Oral and Laryngeal Diadochokinesis Across the Life Span: A Scoping Review of Methods, Reference Data, and Clinical Applications. *J. Speech Lang. Hear. Res.* **2022**, *65*, 574–623. [[CrossRef](#)]
32. Dahl, K.L.; Bolognone, R.K.; Childes, J.M.; Pryor, R.L.; Graville, D.J.; Palmer, A.D. Characteristics associated with communicative participation after total laryngectomy. *J. Commun. Disord.* **2022**, *96*, 106184. [[CrossRef](#)] [[PubMed](#)]
33. Roy, N.; Barkmeier-Kraemer, J.; Eadie, T.; Sivasankar, M.P.; Mehta, D.; Paul, D.; Hillman, R. Evidence-Based Clinical Voice Assessment: A Systematic Review. *Am. J. Speech-Lang. Pathol.* **2013**, *22*, 212–226. [[CrossRef](#)]
34. Rosdi, F.; Salim, S.S.; Mustafa, M.B. An FPN-based classification method for speech intelligibility detection of children with speech impairments. *Soft Comput.* **2019**, *23*, 2391–2408. [[CrossRef](#)]
35. Failla, S.; Al-Zanoon, N.; Smith, N.; Doyle, P.C. The Effects of Contextual Priming and Alaryngeal Speech Mode on Auditory-Perceptual Ratings of Listener Comfort. *J. Voice* **2021**, *35*, 934.e17–934.e23. [[CrossRef](#)]
36. Stipancic, K.L.; Tjaden, K. Minimally Detectable Change of Speech Intelligibility in Speakers With Multiple Sclerosis and Parkinson’s Disease. *J. Speech Lang. Hear. Res.* **2022**, *65*, 1858–1866. [[CrossRef](#)]
37. Malini, S.; Chandrakala, S. Intelligibility assessment of impaired speech using Regularized self-representation based compact supervisors. *Comput. Speech Lang.* **2022**, *74*, 101355. [[CrossRef](#)]
38. Albaqshi, H.; Sagheer, A. Dysarthric Speech Recognition using Convolutional Recurrent Neural Networks. *Int. J. Intell. Eng. Syst.* **2020**, *13*, 384–392. [[CrossRef](#)]
39. Bessell, N.; Gurd, J.M.; Coleman, J. Dissociation between speech modalities in a case of altered accent with unknown origin. *Clin. Linguist. Phon.* **2020**, *34*, 222–241. [[CrossRef](#)] [[PubMed](#)]
40. Moon, A.M.; Kim, H.P.; Cook, S.; Blanchard, R.T.; Haley, K.L.; Jacks, A.; Shafer, J.S.; Fried, M.W. Speech patterns and enunciation for encephalopathy determination—A prospective study of hepatic encephalopathy. *Hepatol. Commun.* **2022**, *6*, 2876–2885. [[CrossRef](#)] [[PubMed](#)]
41. De Cock, E.; Oostra, K.; Bliki, L.; Volkaerts, A.; Hemelsoet, D.; De Herdt, V.; Batens, K. Dysarthria following acute ischemic stroke: Prospective evaluation of characteristics, type and severity. *Int. J. Lang. Commun. Disord.* **2021**, *56*, 549–557. [[CrossRef](#)]
42. Rowe, H.P.; Gutz, S.E.; Maffei, M.F.; Tomanek, K.; Green, J.R. Characterizing Dysarthria Diversity for Automatic Speech Recognition: A Tutorial From the Clinical Perspective. *Front. Comput. Sci.* **2022**, *4*, 770210. [[CrossRef](#)]
43. Stipancic, K.L.; van Brenk, F.; Kain, A.; Wilding, G.; Tjaden, K. Clear Speech Variants: An Investigation of Intelligibility and Speaker Effort in Speakers With Parkinson’s Disease. *Am. J. Speech-Lang. Pathol.* **2022**, *31*, 2789–2805. [[CrossRef](#)]
44. Rosdi, F.; Mustafa, M.B.; Salim, S.S.; Mat Zin, N.A. Automatic speech intelligibility detection for speakers with speech impairments: The identification of significant speech features. *Sains Malays.* **2019**, *48*, 2737–2747. [[CrossRef](#)]
45. Maskeliūnas, R.; Kulikajėvas, A.; Damaševičius, R.; Pribušis, K.; Ulozaitė-Stanienė, N.; Uloza, V. Lightweight Deep Learning Model for Assessment of Substitution Voicing and Speech after Laryngeal Carcinoma Surgery. *Cancers* **2022**, *14*, 2366. [[CrossRef](#)]
46. Kim, H.; Jeon, J.; Han, Y.J.; Joo, Y.; Lee, J.; Lee, S.; Im, S. Convolutional Neural Network Classifies Pathological Voice Change in Laryngeal Cancer with High Accuracy. *J. Clin. Med.* **2020**, *9*, 3415. [[CrossRef](#)]
47. Feng, Y.; Chen, F.; Ma, J.; Wang, L.; Peng, G. Production of Mandarin consonant aspiration and monophthongs in children with Autism Spectrum Disorder. *Clin. Linguist. Phon.* **2022**. [[CrossRef](#)]
48. Vieira, S.T.; Rosa, R.L.; Rodriguez, D.Z. A speech quality classifier based on tree-cnn algorithm that considers network degradations. *J. Commun. Softw. Syst.* **2020**, *16*, 180–187. [[CrossRef](#)]
49. Poncelet, J.; Renkens, V.; Van hamme, H. Low resource end-to-end spoken language understanding with capsule networks. *Comput. Speech Lang.* **2021**, *66*, 101142. [[CrossRef](#)]
50. Cave, R.; Bloch, S. The use of speech recognition technology by people living with amyotrophic lateral sclerosis: A scoping review. *Disabil. Rehabil. Assist. Technol.* **2021**. [[CrossRef](#)] [[PubMed](#)]
51. Schultz, B.G.; Tarigoppula, V.S.A.; Noffs, G.; Rojas, S.; van der Walt, A.; Grayden, D.B.; Vogel, A.P. Automatic speech recognition in neurodegenerative disease. *Int. J. Speech Technol.* **2021**, *24*, 771–779. [[CrossRef](#)]
52. Gupta, S.; Patil, A.T.; Purohit, M.; Parmar, M.; Patel, M.; Patil, H.A.; Guido, R.C. Residual Neural Network precisely quantifies dysarthria severity-level based on short-duration speech segments. *Neural Netw.* **2021**, *139*, 105–117. [[CrossRef](#)]
53. Latha, M.; Shivakumar, M.; Manjula, G.; Hemakumar, M.; Kumar, M.K. Deep Learning-Based Acoustic Feature Representations for Dysarthric Speech Recognition. *SN Comput. Sci.* **2023**, *4*, 272. [[CrossRef](#)]
54. Vishnika Veni, S.; Chandrakala, S. Investigation of DNN-HMM and Lattice Free Maximum Mutual Information Approaches for Impaired Speech Recognition. *IEEE Access* **2021**, *9*, 168840–168849.
55. Chandrakala, S.; Malini, S.; Veni, S.V. Histogram of States Based Assistive System for Speech Impairment Due to Neurological Disorders. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2021**, *29*, 2425–2434. [[CrossRef](#)]
56. Srinivasan, M.; Shanmuganathan, C.; Gupta, S.M.K.; Sikkandar, M.Y. Multi-view representation based speech assisted system for people with neurological disorders. *J. Ambient. Intell. Humaniz. Comput.* **2021**. [[CrossRef](#)]
57. Chandrakala, S.; Malini, S.; Jayalakshmi, S.L. Bag of Models Based Embeddings for Assessment of Neurological Disorders Using Speech Intelligibility. *IEEE Trans. Emerg. Top. Comput.* **2021**, *9*, 1265–1275. [[CrossRef](#)]
58. Fu, J.; Yang, S.; He, F.; He, L.; Li, Y.; Zhang, J.; Xiong, X. Sch-net: A deep learning architecture for automatic detection of schizophrenia. *BioMed. Eng. Online* **2021**, *20*, 75. [[CrossRef](#)] [[PubMed](#)]

59. Marini, M.; Vanello, N.; Fanucci, L. Optimising speaker-dependent feature extraction parameters to improve automatic speech recognition performance for people with dysarthria. *Sensors* **2021**, *21*, 6460. [[CrossRef](#)] [[PubMed](#)]
60. Mathew, L.R.; Gopakumar, K. Evaluation of speech enhancement algorithms applied to electrolaryngeal speech degraded by noise. *Appl. Acoust.* **2021**, *174*, 107771. [[CrossRef](#)]
61. Ishikawa, K.; Boyce, S.; Kelchner, L.; Powell, M.G.; Schieve, H.; de Alarcon, A.; Khosla, S. The Effect of Background Noise on Intelligibility of Dysphonic Speech. *J. Speech Lang. Hear. Res.* **2017**, *60*, 1919–1929. [[CrossRef](#)] [[PubMed](#)]
62. Dhivya, R.; Justin, J. Performance Evaluation of a Speech Enhancement Technique Using Wavelets. In *Proceedings of the International Conference on Soft Computing Systems*; Springer: New Delhi, India, 2015; pp. 637–646. [[CrossRef](#)]
63. Jaiswal, R.K.; Yeduri, S.R.; Cenkeramaddi, L.R. Single-channel speech enhancement using implicit Wiener filter for high-quality speech communication. *Int. J. Speech Technol.* **2022**, *25*, 745–758. [[CrossRef](#)]
64. Pauline, S.H.; Dhanalakshmi, S.; Kumar, R.; Narayanamoorthi, R. Noise reduction in speech signal of Parkinson's Disease (PD) patients using optimal variable stage cascaded adaptive filter configuration. *Biomed. Signal Process. Control* **2022**, *77*, 103802. [[CrossRef](#)]
65. Doi, H.; Toda, T.; Nakamura, K.; Saruwatari, H.; Shikano, K. Alaryngeal Speech Enhancement Based on One-to-Many Eigenvoice Conversion. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2014**, *22*, 172–183. [[CrossRef](#)]
66. Pauline, S.H.; Dhanalakshmi, S. A low-cost automatic switched adaptive filtering technique for denoising impaired speech signals. *Multidimens. Syst. Signal Process.* **2022**, *33*, 1387–1408. [[CrossRef](#)]
67. Pandey, P.; Bhandarkar, S.; Bachher, G.; Lehana, P. Enhancement of alaryngeal speech using spectral subtraction. In *Proceedings of the 2002 14th International Conference on Digital Signal Processing Proceedings*. DSP 2002 (Cat. No.02TH8628), Santorini, Greece, 1–3 July 2002; Volume 2, pp. 591–594. [[CrossRef](#)]
68. Azarnoush, H.; Mir, F.; Agaian, S.; Jamshidi, M.; Shadaram, M. Alaryngeal Speech Enhancement Using Minimum Statistics Approach to Spectral Subtraction. In *Proceedings of the 2007 IEEE International Conference on System of Systems Engineering*, San Antonio, TX, USA, 16–18 April 2007; pp. 1–5. [[CrossRef](#)]
69. Wei, Y.; Li, C.; Li, T.; Zeng, Y. Whispered Speech Enhancement Based on Improved Mel Frequency Scale and Modified Compensated Phase Spectrum. *Circuits Syst. Signal Process.* **2019**, *38*, 5839–5860. [[CrossRef](#)]
70. Mollaei, F.; Shiller, D.M.; Baum, S.R.; Gracco, V.L. The Relationship Between Speech Perceptual Discrimination and Speech Production in Parkinson's Disease. *J. Speech Lang. Hear. Res.* **2019**, *62*, 4256–4268. [[CrossRef](#)]
71. Giri, M.; Rayavarapu, N. Improving the intelligibility of dysarthric speech using a time domain pitch synchronous-based approach. *Int. J. Electr. Comput. Eng.* **2023**, *13*, 4041–4051. [[CrossRef](#)]
72. Ishaq, R.; Shahid, M.; Löfvström, B.; Zapirain, B.G.; Claesson, I. Modulation frequency domain adaptive gain equalizer using convex optimization. In *Proceedings of the 2012 6th International Conference on Signal Processing and Communication Systems*, Gold Coast, QLD, Australia, 12–14 December 2012; pp. 1–5. [[CrossRef](#)]
73. Vijayan, K.; Murty, K.S.R. Prosody Modification Using Allpass Residual of Speech Signals. *Proc. Interspeech* **2016**, *2016*, 1069–1073. [[CrossRef](#)]
74. Bhangale, K.B.; Kothandaraman, M. Survey of Deep Learning Paradigms for Speech Processing. *Wirel. Pers. Commun.* **2022**, *125*, 1913–1949. [[CrossRef](#)]
75. Kobayashi, K.; Toda, T. Implementation of low-latency electrolaryngeal speech enhancement based on multi-task CLDNN. In *Proceedings of the 2020 28th European Signal Processing Conference (EUSIPCO)*, Amsterdam, The Netherlands, 18–21 January 2021; pp. 396–400. [[CrossRef](#)]
76. Saleem, N.; Khattak, M.I.; Alqahtani, S.A.; Jan, A.; Hussain, I.; Khan, M.N.; Dahshan, M. U-Shaped Low-Complexity Type-2 Fuzzy LSTM Neural Network for Speech Enhancement. *IEEE Access* **2023**, *11*, 20814–20826. [[CrossRef](#)]
77. Huq, M. Enhancement of Alaryngeal Speech using Generative Adversarial Network (GAN). In *Proceedings of the 2021 IEEE/ACS 18th International Conference on Computer Systems and Applications (AICCSA)*, Tangier, Morocco, 30 November–3 December 2021; pp. 1–2. [[CrossRef](#)]
78. Pascual, S.; Bonafonte, A.; Serrà, J.; Gonzalez, J.A. Whispered-to-voiced Alaryngeal Speech Conversion with Generative Adversarial Networks. *arXiv* **2018**, arXiv:1808.10687.
79. Pascual, S.; Serrà, J.; Bonafonte, A. Towards Generalized Speech Enhancement with Generative Adversarial Networks. *arXiv* **2019**, arXiv:1904.03418.
80. Amarjounf, M.; Bahja, F.; Di-Martino, J.; Chami, M.; Ibn-Elhaj, E.H. Predicted Phase Using Deep Neural Networks to Enhance Esophageal Speech. In *Lecture Notes on Data Engineering and Communications Technologies*; Springer Nature: Cham, Switzerland, 2023; pp. 68–76. [[CrossRef](#)]
81. Subramanian, A.S.; Wang, X.; Baskar, M.K.; Watanabe, S.; Taniguchi, T.; Tran, D.; Fujita, Y. Speech Enhancement Using End-to-End Speech Recognition Objectives. In *Proceedings of the 2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, 20–23 October 2019; pp. 234–238. [[CrossRef](#)]
82. Muthusamy, H.; Polat, K.; Yaacob, S. Improved Emotion Recognition Using Gaussian Mixture Model and Extreme Learning Machine in Speech and Glottal Signals. *Math. Probl. Eng.* **2015**, *2015*, 394083. [[CrossRef](#)]
83. Li, M.; Wang, L.; Xu, Z.; Cai, D. Mandarin electrolaryngeal voice conversion with combination of Gaussian mixture model and non-negative matrix factorization. In *Proceedings of the 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Kuala Lumpur, Malaysia, 12–15 December 2017; pp. 1360–1363. [[CrossRef](#)]

84. Xu, T.; Feng, K.; Ge, Y.; Zhang, X.; Tao, Z. Identification of vocal nodules and laryngitis by Gauss mixture model. In Proceedings of the 2017 4th International Conference on Systems and Informatics (ICSAI), Hangzhou, China, 11–13 November 2017; pp. 1098–1102. [[CrossRef](#)]
85. Areiza-Laverde, H.J.; Castro-Ospina, A.E.; Peluffo-Ordóñez, D.H. Voice Pathology Detection Using Artificial Neural Networks and Support Vector Machines Powered by a Multicriteria Optimization Algorithm. In *Communications in Computer and Information Science*; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 148–159. [[CrossRef](#)]
86. Das, N.; Chakraborty, S.; Chaki, J.; Padhy, N.; Dey, N. Fundamentals, present and future perspectives of speech enhancement. *Int. J. Speech Technol.* **2020**, *24*, 883–901. [[CrossRef](#)]
87. Lee, S.H.; Kim, M.; Seo, H.G.; Oh, B.M.; Lee, G.; Leigh, J.H. Assessment of Dysarthria Using One-Word Speech Recognition with Hidden Markov Models. *J. Korean Med. Sci.* **2019**, *34*, e108. [[CrossRef](#)]
88. van Sluis, K.E.; van Son, R.J.J.H.; van der Molen, L.; MCGuinness, A.J.; Palme, C.E.; Novakovic, D.; Stone, D.; Natsis, L.; Charters, E.; Jones, K.; et al. Multidimensional evaluation of voice outcomes following total laryngectomy: A prospective multicenter cohort study. *Eur. Arch. Oto-Rhino-Laryngol.* **2020**, *278*, 1209–1222. [[CrossRef](#)]
89. Succo, G.; Peretti, G.; Piazza, C.; Remacle, M.; Eckel, H.E.; Chevalier, D.; Simo, R.; Hantzakos, A.G.; Rizzotto, G.; Lucioni, M.; et al. Open partial horizontal laryngectomies: A proposal for classification by the working committee on nomenclature of the European Laryngological Society. *Eur. Arch. Oto-Rhino-Laryngol.* **2014**, *271*, 2489–2496. [[CrossRef](#)] [[PubMed](#)]
90. Dejonckere, P.H.; Moerman, M.B.J.; Martens, J.P.; Schoentgen, J.; Manfredi, C. Voicing quantification is more relevant than period perturbation in substitution voices: An advanced acoustical study. *Eur. Arch. Oto-Rhino-Laryngol.* **2012**, *269*, 1205–1212. [[CrossRef](#)] [[PubMed](#)]
91. Maskeliūnas, R.; Damaševičius, R.; Kulikajevas, A.; Padervinskis, E.; Pribušis, K.; Uloza, V. A Hybrid U-Lossian Deep Learning Network for Screening and Evaluating Parkinson’s Disease. *Appl. Sci.* **2022**, *12*, 11601. [[CrossRef](#)]
92. Moerman, M.; Martens, J.P.; Dejonckere, P. Multidimensional assessment of strongly irregular voices such as in substitution voicing and spasmodic dysphonia: A compilation of own research. *Logop. Phoniatr. Vocology* **2014**, *40*, 24–29. [[CrossRef](#)]
93. Uloza, V.; Maskeliūnas, R.; Pribušis, K.; Vaitkus, S.; Kulikajevas, A.; Damasevicius, R. An Artificial Intelligence-Based Algorithm for the Assessment of Substitution Voicing. *Appl. Sci.* **2022**, *12*, 9748. [[CrossRef](#)]
94. Campbell, I. Chi-squared and Fisher–Irwin tests of two-by-two tables with small sample recommendations. *Stat. Med.* **2007**, *26*, 3661–3675. [[CrossRef](#)]
95. Schindler, A.; Mozzanica, F.; Ginocchio, D.; Invernizzi, A.; Peri, A.; Ottaviani, F. Voice-related quality of life in patients after total and partial laryngectomy. *Auris Nasus Larynx* **2012**, *39*, 77–83. [[CrossRef](#)]
96. Teruya, N.; Sunagawa, Y.; Toyosato, T.; Yokota, T. Association between Daily Life Difficulties and Acceptance of Disability in Cancer Survivors after Total Laryngectomy: A Cross-Sectional Survey. *Asia-Pac. J. Oncol. Nurs.* **2019**, *6*, 170–176. [[CrossRef](#)]
97. Lin, Z.; Lin, H.; Chen, Y.; Xu, Y.; Chen, X.; Fan, H.; Wu, X.; Ke, X.; Lin, C. Long-term survival trend after primary total laryngectomy for patients with locally advanced laryngeal carcinoma. *J. Cancer* **2021**, *12*, 1220–1230. [[CrossRef](#)]
98. Birkeland, A.C.; Beesley, L.; Bellile, E.; Rosko, A.J.; Hoesli, R.; Chinn, S.B.; Shuman, A.G.; Prince, M.E.; Wolf, G.T.; Bradford, C.R.; et al. Predictors of survival after total laryngectomy for recurrent/persistent laryngeal squamous cell carcinoma. *Head Neck* **2017**, *39*, 2512–2518. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.