INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

Behavioral Analysis of Pathological Speaker Embeddings of Patients During Oncological Treatment of Oral Cancer

Jenthe Thienpondt[1], *Caroline M. Speksnijder*[2], *Kris Demuynck*[1]

[1]IDLab, Department of Electronics and Information Systems, Ghent University - imec, Belgium
[2]Department of Oral and Maxillofacial Surgery and Special Dental Care, University Medical Center Utrecht, Utrecht University, The Netherlands

jenthe.thienpondt@ugent.be, c.m.speksnijder@umcutrecht.nl, kris.demuynck@ugent.be

Abstract

In this paper, we analyze the behavior of speaker embeddings of patients during oral cancer treatment. First, we found that pre- and post-treatment speaker embeddings differ significantly, notifying a substantial change in voice characteristics. However, a partial recovery to pre-operative voice traits is observed after 12 months post-operation. Secondly, the same-speaker similarity at distinct treatment stages is similar to healthy speakers, indicating that the embeddings can capture characterizing features of even severely impaired speech. Finally, a speaker verification analysis signifies a stable false positive rate and variable false negative rate when combining speech samples of different treatment stages. This indicates robustness of the embeddings towards other speakers, while still capturing the changing voice characteristics during treatment. To the best of our knowledge, this is the first analysis of speaker embeddings during oral cancer treatment of patients.

Index Terms: pathological speaker embeddings, oral cancer treatment, speaker recognition

1. Introduction

Oral cancer is a type of cancer that can develop in various locations within the oral cavity, predominantly originating in the tissues of the mouth [1]. It is a serious and potentially life-threatening condition that can cause significant damage to the affected tissues and spread to other parts of the body. Common risk factors for oral cancer include tobacco usage and excessive alcohol consumption [2, 3]. Treatment options for oral cancer typically include surgery, radiation therapy and chemotherapy, which may be used in isolation or in conjunction with each other, depending on the stage and location of the cancer.

In prior research, it is shown that oncological treatment of oral cancer can be accompanied with impaired speech capabilities, including articulation and intelligibility [4, 5, 6]. Subsequent research found reduced speech abilities even after extensive recovery periods up to 12 months after surgical intervention [7]. Another study [8], showed a significant decrease in tongue function during oral cancer treatment, which can potentially be an important contributor to post-intervention speech impairment. Other studies [9, 10] observed a significant decrease in speech recognition transcription accuracy when comparing healthy speakers to a group of patients diagnosed with oral cancer in various treatment stages.

However, to the best of our knowledge, there is no prior research on the behavior of speaker embeddings of patients treated for oral cancer. Speaker embedding similarity, in contrast to conventional intelligibility rating systems, could provide an objective and text-independent measurement of changing voice characteristics without relying on any human perceptual evaluation of pathological speech.

In recent years, speaker verification has gained significant performance increases due to the availability of large and labeled datasets [11, 12], a significant increase in computational power and the advent of specialized deep learning models, including the x-vector architecture [13, 14], ECAPA-TDNN [15] and fwSE-ResNet [16]. Low-dimensional speaker embeddings can be extracted from these models and have shown to capture a wide variety of speaker characteristics, including gender, age, spoken language and emotional state [17, 18, 19].

In this paper, we want to analyze the behavior of speaker embeddings at different stages during oral cancer treatment on multiple properties. First, how do the speaker characteristics, according to the speaker embeddings, evolve between the pre- and post-intervention stages. Subsequently, we want to compare this to previous research results and establish the feasibility of potential usage of speaker embeddings during the oral cancer treatment procedure of a patient. Secondly, assess the intra-session robustness of speaker embeddings of patients based on speech samples recorded at the same session during oral cancer treatment and compare this to a cohort of non-pathological speakers. Finally, perform a speaker verification analysis when combining utterances of several steps in the intervention trajectory of the patients with the goal of analyzing the robustness of the pathological embeddings towards other speakers.

2. Pathological speaker embeddings

Table 1: *Dataset composition of patients with oral cancer.*

	# Male	# Female	# Total
Tumor Stage			
T1 (<2 cm)	5	3	8
T2 (2-4 cm)	11	6	17
T3 (>4 cm)	3	3	6
T4 (metastasis)	15	11	26
Reconstruction Type			
Primary Closure	6	8	14
Local Flap	2	0	2
Free Flap	19	8	27
Bone Flap	7	7	14

The speech samples in the analysis of this paper were collected from 57 Dutch patients with primary oral carcinoma taken at the University Medical Center Utrecht (UMC Utrecht) and the Radboud University Medical Center (Radboudumc) in the Netherlands between January 2007 and August 2009. The study protocol (study ID: NL1200604106) was approved by

10.21437/Interspeech.2023-1868

the Ethics Committees of the UMC Utrecht and Radboudumc. All participants received written information and provided their signed informed consent. The oncological treatment of the patients consists of surgery and subsequent radiotherapy. In addition, samples were also collected from 60 healthy speakers, matched for age and gender, as the control group [20]. Speech samples of patients were taken within 4 weeks before oncological intervention, 4 to 6 weeks after both surgery and radiotherapy and 6 and 12 months after surgery during the recovery phase. The healthy control group has speech samples only taken once. At each sampling session, two speech utterances are collected from the speakers by reading two short, phonetically diverse texts which will be referred to as *text1* and *text2* in this paper, respectively. The texts and recording equipment is kept consistent across all sampling sessions. The average duration of all collected speech samples is 49.6 seconds.

In addition, the tumor stage, as indicated by T of the commonly used TNM cancer staging system [21] of the patients were also collected during the pre-intervention period. The T variable ranges from T1, indicating small tumors, to T4, indicating large tumors which have potentially invaded nearby structures, known as metastasis. Furthermore, the reconstruction type of the oral cancer surgical procedure is also collected, existing of primary closure, free flap, local flap, and bone flap reconstruction. Primary closure refers to the immediate closure of the incision after the removal of cancerous tissue. Local flap reconstruction uses adjacent oral cavity tissue to reconstruct the affected area after tumor removal, while free flap reconstruction uses tissue from another body part. Bone flap reconstruction is used to rebuild bone structures inside the oral cavity after removal of the cancer tumors. The composition of the speaker characteristics of the patients in the dataset is given in Table 1.

The speaker embeddings are extracted from the state-of-the-art speaker verification fwSE-ResNet34 model presented in [16]. This architecture extends the popular ResNet [22] backbone with a speech-adapted version of Squeeze-Excitation (SE) [23] and incorporates positional encodings to extend the spatial invariance of the 2D convolutional kernels with a notion of frequency positional information. The model is optimized using the Additive Angular Margin (AAM) softmax loss function [24], resulting in the cosine distance being the similarity metric between speaker embeddings. More information about the architecture and training procedure can be found in the accompanying paper [16]. We note that this includes using the same training set, which solely exists of the development part of VoxCeleb2 [12], with no form of subsequent domain adaptation to pathological speech.

3. Pathological speaker analysis

It is shown that various functions related to the oral cavity are impacted by surgical and radiotherapy interventions, including the masticatory, swallowing and speech capabilities [25, 6, 8]. To analyze the evolution of the speaker identifying characteristics of patients undergoing oral oncological treatment, we calculate the cosine similarities speaker-wise between the pre-operative *text1* embedding and all *text2* embeddings at different stages in the treatment trajectory. We also calculate the cosine similarities between the *text1* and *text2* embeddings of the healthy speakers to be compared to the pre-operative embedding behavior of the patients.

Figure 1 depicts a box plot describing the evolution of the speaker similarity relative to the pre-operative speaker embeddings. We observe no significant difference between the pre-operative speaker similarity of the pathological group in comparison to the healthy set of speakers. While pre-operative speech impairment is usually limited for patients diagnosed with oral cancer in comparison to the post-intervention condition [26], it is encouraging to observe similar behavior of pre-operative pathological speakers and the healthy control group. Section 3.3 analyzes the intra-session robustness of the speaker embeddings in more detail.

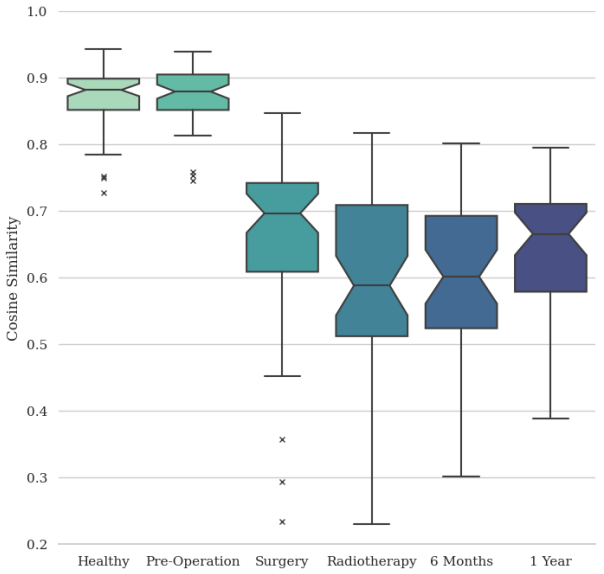

Figure 1: *Tukey-style box plot depicting the evolution of pre-operative speaker embedding similarity of patients (n=57) during oral cancer treatment. Speaker similarity of a healthy control group (n=60) is included as reference. Notch width indicates the 95% confidence interval of the median.*

A significant decrease in pre-operative speaker similarity is observed after surgical treatment of the patients. It is previously shown that surgical intervention in oral cancer treatment has a significant negative impact on a wide variety of oral function abilities, including self-reported speech capability [27, 28]. Those findings are reinforced by observing a comparable degradation between the pre-operative and post-operative speaker embedding similarity, which provides an objective and robust measurement of changing voice characteristics.

Radiotherapy during oral oncological treatment can potentially impact important tissues related to speech production [29]. However, the cumulative effect on oral function of post-operative radiotherapy strongly depends on variables such as tumor location, tumor stage and reconstruction type [27]. In our results, an additional significant change in voice characteristics is discerned after the post-operative radiotherapy stage in the treatment trajectory. We also observe a substantial increase in variability between pre-operative and post-radiotherapy speaker similarity, suggesting the final extent of change in voice characteristics is highly dependent on some underlying variables.

Both an increased pre-operative speaker similarity and decreased variability is noted after the 6-month recovery period, relative to the post-radiotherapy stage, with a similar trend in the following 6 months. This indicates that voice characteristics tend to return to the pre-operative state to a certain extent for at least a 1-year period post-intervention.

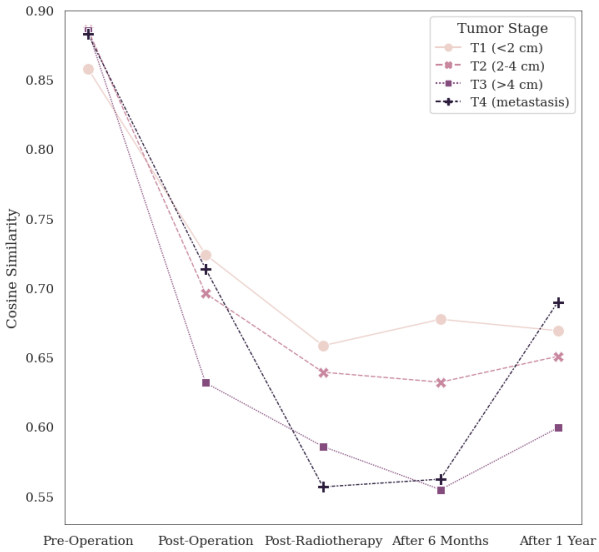

Figure 2: *Effect of tumor stage, as measured by T of the TNM cancer staging model, on the evolution of pre-operative voice similarity of patients during oral cancer treatment.*

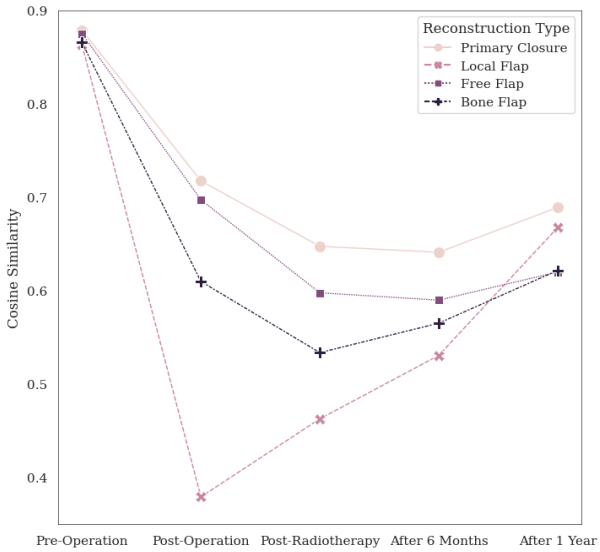

Figure 3: *Effect of surgical reconstruction type on the evolution of pre-operative voice similarity of patients during oral cancer treatment.*

3.1. Tumor stage impact on voice characteristics

Figure 2 depicts the change in pre-operative voice characteristics for each subgroup of patients based on tumor stage determined before intervention. The figure shows the mean cosine similarity between the pre-operative *text1* and pre- and post-operative *text2* embeddings for each subgroup. The number of speakers in each group is given in Table 1.

We notice an inversely proportional relationship between the tumor size and the pre-operative speaker similarity at the post-intervention stages. This corroborates previous research which suggests that late-stage tumors were associated with poorer post-operative speech outcomes, including reduced speech intelligibility and decreased vocal quality [30]. Notably, this is accompanied with a more pronounced recovery towards pre-operative speaker characteristics in the T3 and T4 groups after the 1-year post-intervention period. This suggests that the additional severity of post-radiotherapy changes in speaker characteristics in the late-stage tumor groups is partially or even completely offset after sufficient recovery time.

3.2. Reconstruction type impact on voice characteristics

Likewise, Figure 3 shows the evolution of the pre-operative mean speaker similarity according to the type of reconstructive surgery performed. We observe that primary closure has the least significant impact on post-intervention voice characteristics in comparison to flap-based reconstruction. This supports previous research in which patients treated with primary closure were rated higher in speech intelligibility [31]. Notable is the significantly more severe change of voice characteristics of patients undergoing restorative local flap surgery in comparison to free flap surgery. This can possibly be attributed to the removal of tissue from the oral cavity during local flap surgery, as opposed to tissue removal from other parts of the body in free flap surgery. The removal of tissue in the oral cavity can potentially devise an additional degree of voice transformation in the patient in the case of local flap restoration. However, we note that the number of local flap surgeries in our dataset is limited.

3.3. Intra-session robustness of pathological embeddings

State-of-the-art speaker embeddings have shown to robustly capture speaker characteristics in a variety of challenging conditions, including severe background noise, short sampling duration and language switching [32]. However, it is an open question how well these embeddings can identify speakers who have had severe medical intervention in the oral cavity region. Surgery related to oral cancer treatment can have a severe impact on the structural composition of the vocal tract, which could potentially both limit or enhance the identifying characteristics captured by the speaker embeddings. In this section, we analyze the intra-session robustness of the speaker embeddings at all stages during oral cancer treatment.

To establish the intra-session robustness of the speaker embeddings, we calculate the cosine similarity between the *text1* and *text2* embedding of each patient at all sampling sessions during the treatment trajectory. The session-wise mean and standard deviation of the same-speaker cosine similarities is shown in Figure 4. As a reference, the mean similarity between the embeddings from the same speakers in the healthy group is indicated by the dotted line. For comparison, we also plotted the mean and standard deviation of the speaker-wise similarities between the pre-operative *text1* and post-intervention *text2* embeddings.

We can observe that the mean intra-session similarity is very consistent during the complete oral cancer treatment trajectory, even slightly exceeding the healthy control group. This indicates that the speaker embeddings are able to capture robust and distinguishing voice characteristics of speakers, even after substantial oncological intervention in the oral cavity, given the changed voice characteristics are temporally stable. This is notable due to the training set of the speaker embedding extractor not containing any comparable pathological speakers. This implies no domain-specific adaption of the training procedure of the speaker embedding extractor is needed, which greatly alleviates the potential medical usage of speaker embeddings in oral cancer treatment.

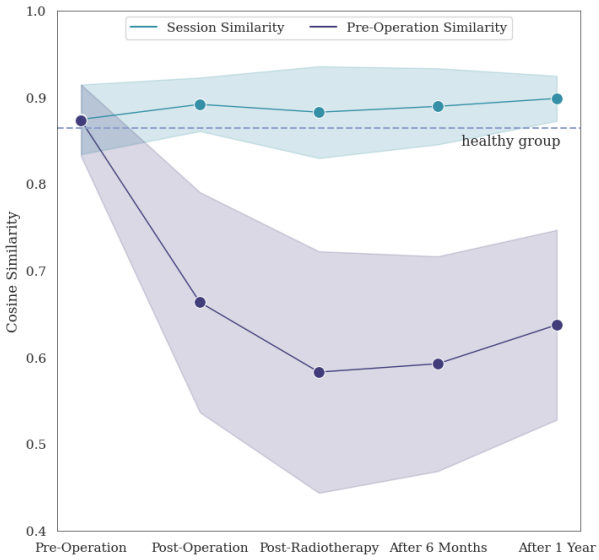

Figure 4: *Intra-session and inter-session (relative to the pre-operative session) same-speaker similarity of patients during oral cancer treatment. The dotted line indicates the mean same-speaker similarity of the healthy control group.*

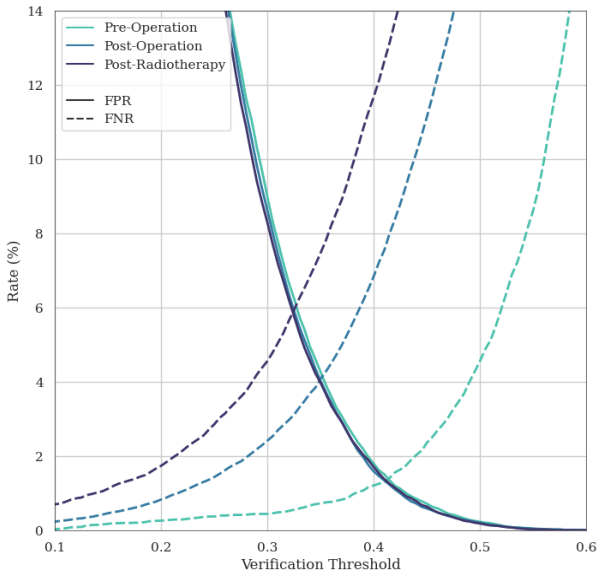

Figure 5: *False positive rates (FPR) and false negative rates (FNR) of speaker verification trials consisting of pre-operative, pre-operative combined with post-operative and pre-operative combined with post-radiotherapy speech samples.*

3.4. Pathological speaker verification analysis

In this section we want to analyze the behavior of pathological speaker embeddings in a speaker verification setting. Speaker verification attempts to solve the task if two utterances are spoken by the same person. We create three groups of speaker verification trials based on speech samples from patients: pre-operative, pre-operative combined with post-operative and pre-operative combined with post-radiotherapy utterances. To increase the number of trials, we create consecutive, non-overlapping crops of 5 seconds of each utterance and subsequently extract the speaker embeddings as described in Section 2. Each trial consists of a *text1* embedding paired with a *text2* embedding for text-independency and we balance the amount of positive and negative trials. Results are reported using the equal error rate (EER) and a breakdown of the false positive rate (FPR) and false negative rate (FNR).

Table 2: *Speaker verification results of oral cancer patients. FPR and FNR are based on a threshold value of 0.35.*

	EER (%)	**FPR (%)**	**FNR (%)**
Pre-operation	1.39	4.26	0.73
Post-operation	4.06	4.03	4.08
Post-radiotherapy	5.88	3.97	7.48

As shown in Table 2, the overall EER sharply increases by the subsequent addition of post-operative and post-radiotherapy samples. However, as Figure 5 indicates, the FPR of all trial groups remains almost identical, independent of the chosen speaker verification threshold. The degradation of EER can exclusively be attributed by an increase in FNR in the groups combining pre-operative and post-intervention embeddings. The implications of a stable FPR and variable FNR are desirable from an oral cancer treatment viewpoint. A stable FPR signifies a robust behavior of the speaker embeddings towards other speakers, while simultaneously still being able to capture the change in voice characteristics of the same speaker during the treatment trajectory.

4. Future work

As shown in this paper, the use of speaker embeddings has the potential to improve our understanding of changing voice characteristics during oral cancer treatment. Using speaker embeddings to analyze individual treatment trajectories proves viable due to a combination of intra-session robustness, objective and text-independent metrics for changing voice characteristics and no reliance on human perceptual evaluation in the process. In future work, we will attempt to investigate the feasibility of using speaker embeddings to identify potential complications or challenges that may arise during the recovery process.

5. Conclusion

In this paper, we analyzed the behavior of speaker embeddings of patients diagnosed with oral cancer at different stages during oncological treatment. First, we found that pre-operative and post-intervention speaker similarity significantly diminishes. However, we observe an evolution of the voice characteristics towards the pre-operative stage in the following 12-month post-operative period. Secondly, we establish the intra-session robustness of current state-of-the-art speaker embeddings on speakers with oral cancer treatment. This indicates that the embeddings can successfully capture pathological speaker characteristics, given the pathology is temporally stable. Finally, we observe a stable false positive rate and variable false negative rate in a speaker verification analysis when speech samples are used from different stages in oral cancer treatment. This signifies a stable behavior of the embeddings towards other speakers while still being able to capture the change in voice characteristics during oral oncological treatment.

6. References

[1] D. M. Parkin, F. Bray, J. Ferlay, and P. Pisani, "Estimating the world cancer burden: Globocan 2000," *International Journal of Cancer*, vol. 94, no. 2, pp. 153–156, 2001.

[2] M. de Boer, R. Sanderson, R. Damhuis, C. Meeuwis, and P. Knegt, "The effects of alcohol and smoking upon the age, anatomic sites and stage in the development of cancer of the oral cavity and oropharynx in females in the south west netherlands." in *Eur Arch Otorhinolaryngol.*, 1997, pp. 177–179.

[3] D. Morse, W. Psoter, D. Cleveland, D. Cohen, M. Mohit-Tabatabai, D. Kosis, and E. Eisenberg, "Smoking and drinking in relation to oral cancer and oral epithelial dysplasia," *Cancer causes & control : CCC*, vol. 18, pp. 919–29, 2007.

[4] J. Hufnagle, P. Pullon, and K. Hufnagle, "Speech considerations in oral surgery: Part i. speech physiology," *Oral Surgery, Oral Medicine, Oral Pathology*, vol. 46, no. 3, pp. 349–353, 1978.

[5] J. Hufnagle, P. A. Pullon, and K. Hufnagle, "Speech considerations in oral surgery. part ii. speech characteristics of patients following surgery for oral malignancies." *Oral surgery, oral medicine, and oral pathology*, vol. 46 3, pp. 354–61, 1978.

[6] B. R. Pauloski, A. W. Rademaker, J. A. Logemann, and L. A. Colangelo, "Speech and swallowing in irradiated and nonirradiated postsurgical oral cancer patients," *Otolaryngology–Head and Neck Surgery*, vol. 118, no. 5, pp. 616–624, 1998.

[7] P. A. Borggreven, I. V. de Leeuw, J. A. Langendijk, P. Doornaert, M. N. Koster, R. de Bree, and C. R. Leemans, "Speech outcome after surgical treatment for oral and oropharyngeal cancer: A longitudinal assessment of patients reconstructed by a microvascular flap," *Head & Neck*, vol. 27, no. 9, pp. 785–793, 2005.

[8] C. Speksnijder, A. Bilt, H. van der Glas, R. Koole, and M. A. Merkx, "Tongue function in patients treated for malignancies in tongue and/or floor of mouth; a one year prospective study," *International journal of oral and maxillofacial surgery*, vol. 40, pp. 1388–94, 2011.

[9] M. Windrich, A. Maier, R. Kohler, E. Noeth, E. Nkenke, U. Eysholdt, and M. Schuster, "Automatic quantification of speech intelligibility of adults with oral squamous cell carcinoma," *Folia phoniatrica et logopaedica : official organ of the International Association of Logopedics and Phoniatrics (IALP)*, vol. 60, pp. 151–6, 2008.

[10] B. M. Halpern, S. Feng, R. van Son, M. van den Brekel, and O. Scharenborg, "Low-resource automatic speech recognition and error analyses of oral cancer speech," *Speech Communication*, vol. 141, pp. 14–27, 2022.

[11] A. Nagrani, J. S. Chung, and A. Zisserman, "VoxCeleb: A large-scale speaker identification dataset," in *Proc. INTERSPEECH 2017 – 18th Annual Conference of the International Speech Communication Association*, 2017, pp. 2616–2620.

[12] J. S. Chung, A. Nagrani, and A. Zisserman, "VoxCeleb2: Deep speaker recognition," in *Proc. INTERSPEECH 2018 – 19th Annual Conference of the International Speech Communication Association*, 2018, pp. 1086–1090.

[13] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, and S. Khudanpur, "X-vectors: Robust dnn embeddings for speaker recognition," in *ICASSP 2018*, 2018, pp. 5329–5333.

[14] D. Snyder, D. Garcia-Romero, G. Sell, A. McCree, D. Povey, and S. Khudanpur, "Speaker recognition for multi-speaker conversations using x-vectors," in *ICASSP 2019*, 2019, pp. 5796–5800.

[15] B. Desplanques, J. Thienpondt, and K. Demuynck, "ECAPA-TDNN: Emphasized channel attention, propagation and aggregation in TDNN based speaker verification," in *Proc. INTERSPEECH 2020 – 21st Annual Conference of the International Speech Communication Association*, 2020, pp. 3830–3834.

[16] J. Thienpondt, B. Desplanques, and K. Demuynck, "Integrating frequency translational invariance in TDNNs and frequency positional information in 2D ResNets to enhance speaker verification," in *Proc. INTERSPEECH 2021 – 22rd Annual Conference of the International Speech Communication Association*, 2021, pp. 2302–2306.

[17] R. Pappagari, T. Wang, J. Villalba, N. Chen, and N. Dehak, "X-vectors meet emotions: A study on dependencies between emotion and speaker recognition," *ICASSP 2020*, pp. 7169–7173, 2020.

[18] D. Raj, D. Snyder, D. Povey, and S. Khudanpur, "Probing the information encoded in x-vectors," in *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2019, pp. 726–733.

[19] D. Kwasny and D. Hemmerling, "Gender and age estimation methods based on speech using deep neural networks," *Sensors*, vol. 21, no. 14, 2021.

[20] C. Speksnijder, J. Abbink, H. van der Glas, N. Janssen, and A. Bilt, "Mixing ability test compared to a comminution test in persons with normal and compromised masticatory performance," *European journal of oral sciences*, vol. 117, pp. 580–6, 2009.

[21] S. G. Patel and J. P. Shah, "TNM staging of cancers of the head and neck: Striving for uniformity among diversity," *CA: A Cancer Journal for Clinicians*, vol. 55, no. 4, pp. 242–258, 2005.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE/CVF CVPR*, 2016, pp. 770–778.

[23] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.

[24] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4685–4694.

[25] C. Speksnijder, A. Bilt, J. Abbink, M. A. Merkx, and R. Koole, "Mastication in patients treated for malignancies in tongue and/or floor of mouth: A 1-year prospective study," *Head & Neck*, vol. 33, pp. 1013 – 1020, 2011.

[26] G. Saravanan, V. Ranganathan, A. Gandhi, and V. Jaya, "Speech outcome in oral cancer patients – pre- and post-operative evaluation: A cross-sectional study," *Indian Journal of Palliative Care*, vol. 22, pp. 499–503, 2016.

[27] M.-M. Suarez-Cunqueiro, A. Schramm, R. Schoen, J. Seoane-Lestón, X.-L. Otero-Cepeda, K.-H. Bormann, H. Kokemueller, M. Metzger, P. Diz-Dios, and N.-C. Gellrich, "Speech and swallowing impairment after treatment for oral and oropharyngeal cancer," *Archives of Otolaryngology–Head & Neck Surgery*, vol. 134, no. 12, pp. 1299–1304, 2008.

[28] C. Speksnijder, H. van der Glas, A. van der Bilt, R. van Es, E. van der Rijt, and R. Koole, "Oral function after oncological intervention in the oral cavity: A retrospective study," *Journal of Oral and Maxillofacial Surgery*, vol. 68, no. 6, pp. 1231–1237, 2010.

[29] S. Huang and B. O'Sullivan, "Oral cancer: Current role of radiotherapy and chemotherapy," *Medicina oral, patologia oral y cirugia bucal*, vol. 18, 2013.

[30] L. Thomas, T. Jones, S. Tandon, P. Carding, D. Lowe, and S. Rogers, "Speech and voice outcomes in oropharyngeal cancer and evaluation of the university of washington quality of life speech domain," *Clinical otolaryngology : official journal of ENT-UK ; official journal of Netherlands Society for Oto-Rhino-Laryngology & Cervico-Facial Surgery*, vol. 34, pp. 34–42, 2009.

[31] F. M. S. McConnel, B. R. Pauloski, J. A. Logemann, A. W. Rademaker, L. Colangelo, D. Shedd, W. Carroll, J. Lewin, and J. Johnson, "Functional results of primary closure vs flaps in oropharyngeal reconstruction: A prospective study of speech and swallowing," *Archives of Otolaryngology–Head & Neck Surgery*, vol. 124, no. 6, pp. 625–630, 1998.

[32] J. Thienpondt, B. Desplanques, and K. Demuynck, "Tackling the score shift in cross-lingual speaker verification by exploiting language information," in *ICASSP 2022*, 2022, pp. 7187–7191.