



Evaluation of different antenna types and positions in a stepped frequency continuous-wave radar-based silent speech interface

João Vítor Possamai de Menezes¹, Pouriya Amini Digehsara¹, Christoph Wagner¹, Marco Mütze², Michael Bärhold², Petr Schaffer², Dirk Plettemeier², Peter Birkholz¹

¹Institute of Acoustics and Speech Communication, Technische Universität Dresden

²Institute of Communication Technology, Technische Universität Dresden

joao.vitor.possamai.de.menezes@tu-dresden.de

Abstract

Silent speech interfaces (SSIs) are subject of growing interest, as they can enable speech communication even in the absence of the acoustic signal. Among sensing techniques used in SSIs, radar sensing has many desirable characteristics, such as non-invasiveness and comfort. Although promising results have been achieved with radar-based SSIs, some of its crucial parameters are yet to be investigated, e.g., the optimal type and position of the antennas. To fill this gap, this study investigated the performance of a radar-based SSI with 3 antenna types attached to 3 positions on the speaker's cheek (9 setups). A corpus of 25 phonemes uttered under co-articulation effects was recorded with the 9 setups by 2 native German speakers and then classified with respect to the phonemes. A linear mixed-effect model was fitted to the resulting recognition rates and likelihood ratio tests showed significance for the effects of antenna type and position. The two monopole-type antennas performed better than the Vivaldi-type antenna ($2.7\% \pm 2.8\%$ and $6.2\% \pm 3.0\%$ improvement), and the two positions closer to the speaker's lips performed better than the most distant position (decrease of $2.8\% \pm 0.9\%$). This provides more solid foundation for the development of this type of SSI.

Index Terms: silent speech interfaces, continuous-wave radar, monopole antenna, Vivaldi antenna, linear mixed-effect models

1. Introduction

The advent of silent speech interfaces (SSIs) opened new possibilities for speech communication when its acoustic signal is totally or partially unavailable. This is due to the ability of SSIs to sense speech-related bio-signals such as muscle [1, 2], brain [3, 4] or articulatory [5, 6] activity and use them as input to speech recognition and/or synthesis systems. Some of the goals of SSIs are to allow speech impaired people to communicate more fluidly and to enable speech communication in severely noisy environments. Current challenges of SSIs are improving stability and portability while also being convenient and non-invasive to use. For reviews on SSI research, see [7, 8, 9].

Some of the most developed sensing technologies in SSIs are surface electromyography (sEMG) [1, 2] and permanent magnet articulography (PMA) [10, 5]. A device composed of 11 face- and neck-worn sEMG sensors was the basis for a SSI able to achieve 91.1% subject-specific word recognition rate with a large vocabulary [1]. The SSI developed by [2] considered channel correlation of multi-channel sEMG, converting signals from time- to frequency-domain and achieving 90% word recognition rate with a 10-word corpus. Direct synthesis in real time was performed by [10, 5] with a PMA-based SSI, whose sensors were attached to the speaker's lips and tongue,

achieving an intelligibility rate of 92%. Another sensing technology for SSIs is the electro-optical stomatography (EOS), which measures jaw, lip and tongue movement using electrical contact and optical sensors mounted on a pseudopalate, worn inside the speaker's mouth. Subject-specific and -independent word recognition rates of 97% and 56%, respectively, were achieved with a 30-word corpus by [6]. Despite impressive results, these systems may still be improved, e.g., by reducing the number of sensors to improve stability, portability and convenience of use (11 in [1] and 6 in [2, 10, 5]), and are still invasive, in the case of PMA- and EOS-based SSIs.

As an alternative, *radar-based* SSIs have the potential to overcome difficulties faced by other SSIs. The sensing of speech-related signals by radar sensors was first presented by [11], but the development of SSIs based upon that began later either with antennas a few centimeters away from the speaker [12, 13] or attached to the speaker's face [14, 15, 16]. Whereas contactless radar-based SSIs sense speech motion from the radar signal reflected from the speaker's face, SSIs with antennas attached to the speaker's face capture transmission and reflection spectra from the upper vocal tract. Radar-based SSIs have shown potential for practical applications, being comfortable and non-invasive to use and presenting relatively high recognition rates. Using contactless antennas, patterns were found in I/Q radar signals of words and sentences [12], whereas a 88% subject-specific and 81.80% session and subject independent word recognition rates were achieved with a 13-word corpus [13]. Using antennas on the speaker's face, [14] achieved 93% subject-specific recognition rate with a 25-phoneme corpus, whereas [16] achieved 99% and 89% subject-specific recognition rates with a 50-word corpus in multi-session and inter-session settings, respectively. Despite the encouraging results shown so far, challenges like system portability and variability of speakers and sessions have not yet been completely overcome and understood.

Additionally, the SSI developed by [14, 16] has parameters whose effects have not yet been investigated, e.g., the optimal type and position of the antennas. These studies used one single antenna type, without an empirical basis to determine its type and position. In this study we wanted to investigate if the recording setup used by [14, 16] can be improved by either using another type of antenna and/or positioning it differently on the speaker's face. For that, we recorded the same corpus as [14] with the same hardware as [16], but with different setups, varying antenna types and positions. We then performed phoneme recognition tasks and analysed the significance of each parameter with contingency tables and linear mixed effect models using recognition rates as the dependent variable.

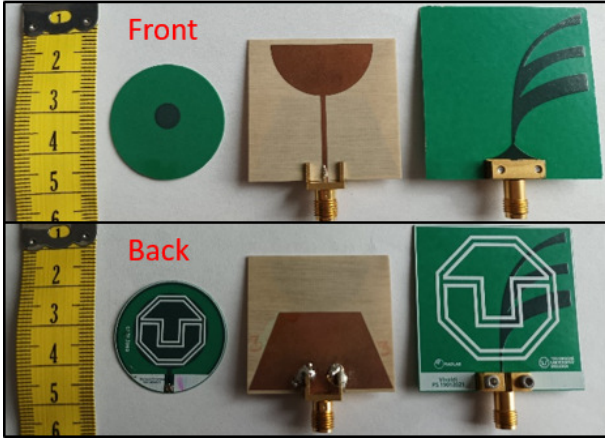


Figure 1: Front and back of the used antennas. A tape with centimeter markings serves as a size reference. Left: disc-shaped monopole antenna with a guard ring. Center: half-disc monopole antenna. Right: antipodal Vivaldi antenna.

2. Experiment design

The experiment designed to compare the performance of our radar-based SSI used three different antenna types: a disc-shaped monopole antenna with a guard ring (MD), a half-disc monopole antenna (MH), and an antipodal Vivaldi antenna (AV), as shown in Figure 1.

The AV antenna is exactly the same as in [16] and an updated version of the antenna used in [14], with a solid substrate. It served as the baseline antenna type and is presented in more detail in [17]. The MH antenna is exactly the same as in [18]. Monopole antennas are usually used for on-body applications [18, 19] and relevant parameters are their height, ideally as small as possible, and the curvature of the surface they are attached to, ideally flat [20]. To improve these parameters, the MD antenna was used, as it is smaller and more likely to have a flat contact surface with the skin compared to the MH antenna.

To solely vary antenna types may not yield the fairest comparison. Since the investigated antennas have different sizes and cover different areas of the upper vocal tract with their emission characteristics depending on where they are fixated, we varied antenna position in the experiment, as shown in Figure 2.

In summary, we varied between three antenna types and three antenna positions. In all positions, the center point of the antenna was aligned with the line between the speaker’s lips. The positions differ with regard to the distance between the edge of the antenna and the corner of the lips (0.5 cm in A, 1.5 cm in B, and 2.5 cm in C). The three chosen positions are appropriate to investigate the antenna behavior horizontally along the upper vocal tract, assessing the movement of articulators, mainly the tongue, along this axis.

The corpus used in the experiment is detailed in Table 1 and is the same as in [14]. It consists of logatomes of 15 vowels, each in the context of 10 consonants, and of 10 consonants, each in the context of 8 vowels. The total number of logatomes was, thus, $15 \times 10 + 10 \times 8 = 230$. The corpus was recorded once for each of the nine combinations of antenna type and position (setups). This procedure was carried out for two native German speakers, in two different sessions for each. The grand total of recorded tokens was $8280 (2 \text{ speakers} \times 2 \text{ sessions} \times 9 \text{ setups} \times 230 \text{ tokens} = 8280)$.

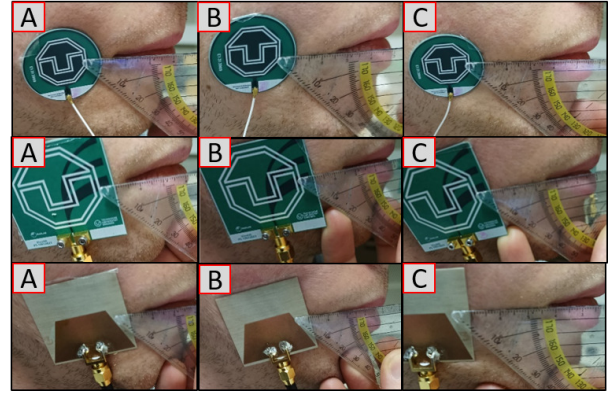


Figure 2: The nine used combinations of antenna type and position. Top row: MD antenna. Middle row: AV antenna. Bottom row: MH antenna. Positions are indicated by A, B, and C.

Table 1: Logatomes composing the corpus, which were produced for all context consonants $C \in \{b, d, g, l, r, f, s, j, m, n\}$ and all context vowels $V \in \{a:, e:, i:, o:, u:, \varepsilon:, \emptyset:, y:/\}$. The target phonemes are underlined and all logatomes were preceded by the German article “Eine” during the recording.

Tense vowels	Lax vowels	Consonants
/C <u>a:</u> də/	/C <u>i</u> tə/	/V <u>b</u> V/
/C <u>e:</u> də/	/C <u>ɛ</u> tə/	/V <u>d</u> V/
/C <u>i:</u> də/	/C <u>a</u> tə/	/V <u>g</u> V/
/C <u>o:</u> də/	/C <u>ɔ</u> tə/	/V <u>l</u> V/
/C <u>u:</u> də/	/C <u>u</u> tə/	/V <u>r</u> V/
/C <u>ɛ:</u> də/	/C <u>y</u> tə/	/V <u>f</u> V/
/C <u>ø:</u> də/	/C <u>œ</u> tə/	/V <u>s</u> V/
/C <u>y:</u> də/		/V <u>j</u> V/
		/V <u>m</u> V/
		/V <u>n</u> V/

3. Data processing and analysis

Recordings took place in an acoustically treated room, as shown in Figure 3. The data flow can be described as follows: the left-cheek antenna emits a stepped frequency continuous wave signal (128 linearly spaced steps between 1 GHz and 6 GHz), whereas the right-cheek antenna receives this signal after transmission through the upper vocal tract. The radio-frequency (RF) signal flows from this antenna into the hardware. It is amplified, then goes through a downmixer stage, which converts it to a fixed intermediate frequency (IF) of 1 MHz, and is finally low-pass filtered (cut-off frequency of 1.407 MHz) before being sampled with an analog-to-digital converter. Finally, this processed received signal is sent to the micro-processing unit, which sends it to a PC (computer) via a USB 2.0 connection.

The resulting raw radar data are vectors of complex numbers sampled at the rate of 100 Hz. Each vector represents the transmission spectrum from the left cheek to the right cheek of the speaker during the utterance of the corpus. The transmission spectrum was obtained by the division of the signal received by the right-cheek antenna by the signal sent by the left-cheek antenna. This procedure is described in greater detail in [16].

To extract the target phonemes, recorded data were segmented with Praat [21] using the audio recorded simultane-

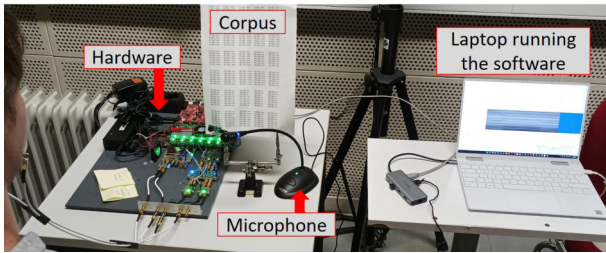


Figure 3: Recording setup of the experiment. The speaker is located on the left, with the antennas attached to his cheek.

Table 2: Grid search values used in the hyperparameter optimization of the SVM and MLP classifiers. Solver (lbfgs) and maximum number of iterations (750) were fixed for MLP.

Classifier	Hyper-param.	Values
SVM	C	{1, 10, 50, 100}
	gamma	{ 10^{-4} , 10^{-3} , 10^{-2} , 10^{-1} }
MLP	activation	{tanh, relu}
	hidden layer size	{100, 50}
	learning rate	{constant, adaptive}
	alpha	{0.0001, 0.05}

ously with the radar data. Only the middle frame of each phoneme was extracted and used further, aiming at the frame where the articulation is most stable, with the least influence of co-articulation from adjacent phonemes. As a result, each target phoneme was represented by a 128-dimensional vector of complex numbers.

To objectively measure how well our SSI performed with each antenna type and position setup, we performed a classification task with three different classifiers (Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), and Multi-Layer Perceptron (MLP)) whose goal was to classify the 25 phonemes of the corpus. The evaluation was carried out with a nested cross-validation (CV), with the inner CV optimizing the hyperparameters of SVM and MLP by grid-search (see Table 2), and the outer CV running the classifiers with optimal hyperparameters. This analysis was performed separately for each speaker and each session. The inner CV was 6-fold and the outer CV was 8-fold, as those were the highest number of folds that allowed one token of each class in each fold. This procedure was performed using the scikit-learn library [22].

As input to the classifiers, we used four feature sets based on transformations of the raw radar data: spectral magnitude, phase, impulse response and a concatenation of spectral magnitude and phase. Feature sets were also standardized (mean value set to 0 and variance set to 1) individually, i.e., when magnitude and phase were concatenated, one standardization took place for each feature. Figures 4 and 5 depict examples of the magnitude and impulse response features.

The significance of the parameters in the experiment can be assessed by analyzing the phoneme recognition rates. This analysis was composed of contingency tables and linear mixed effect models (LMEM). The contingency tables allow a visual overview of the results, with accuracy rates pooled within variables. The LMEM fits linear models to the data, taking into account fixed and random effects, while also being subject to statistical tests.

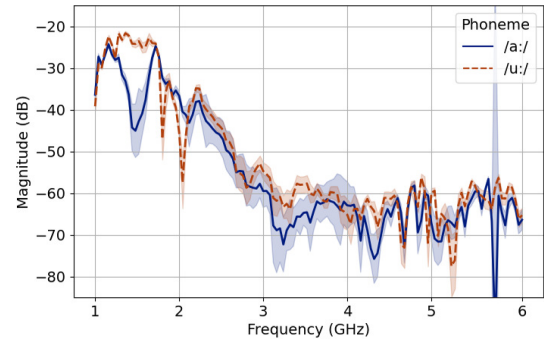


Figure 4: Spectral magnitude of the raw radar data of tense vowels /a:/ and /u:/ under different contexts with antenna type MD and position A. The shades represent the standard deviation across repetitions.

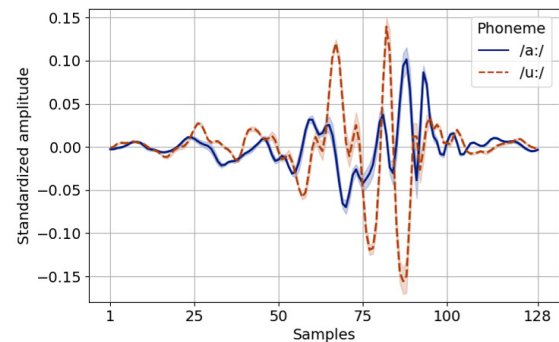


Figure 5: Impulse response of the raw radar data of tense vowels /a:/ and /u:/ under different contexts with antenna type MD and position A. The shades represent the standard deviation across repetitions.

4. Results and discussion

An overview of the obtained recognition rates is shown in Figure 6. The contingency tables obtained by pooling the recognition rates into the variables of interest of the experiment are shown in Figure 7, where some trends can be visually identified. Speaker 2 achieved higher recognition rates than speaker 1 across all variables. Similarly, SVM and MLP classifiers achieved higher recognition rates than LDA across all variables. Feature sets composed of the spectral magnitude and of its concatenation with phase also showed higher recognition rates than other features sets across all variables. Antenna types also achieved different recognition rates across variables, with the AV antenna having the lowest. The variation of antenna position also affected the recognition rates, but different antennas presented different trends.

Each variable seems to have an effect on the recognition rates obtained by our radar-based SSI, but to ensure the significance of these effects, we fitted a LMEM to our data, with recognition rate as dependent variable and speaker, classifier, feature set, antenna type and position as independent variables.

We used R [23] and lme4 [24] to fit the LMEM and analyze how the recognition rate is affected by antenna type and position. As fixed effects we considered antenna type and position, with an interaction term between them. This choice was made, given that the size of the antennas varies with their type, justifying the assumption that the same displacement may have

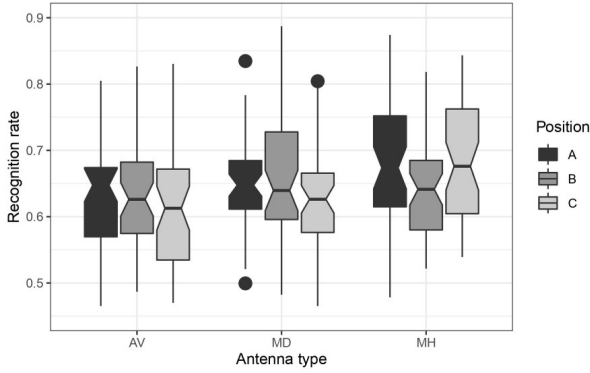


Figure 6: Boxplots of the recognition rates obtained by the classification task. The boxplots are here grouped by antenna type and position.

Phoneme recognition rates		Classifier			Feature Set			
		LDA	SVM	MLP	MAG	PHA	M&P	IMP
Subject	1	55,28%	62,79%	63,28%	65,00%	55,25%	63,91%	57,63%
	2	64,41%	72,90%	73,51%	75,80%	66,50%	75,90%	62,89%
		Classifier	LDA	62,76%	56,86%	62,93%	56,83%	
			SVM	74,06%	62,75%	73,43%	61,13%	
			MLP	74,38%	63,02%	73,37%	62,82%	

Phoneme recognition rates		Antenna			Position		
		MD	MH	AV	A	B	C
Subject	1	61,05%	63,94%	56,35%	60,90%	60,12%	60,33%
	2	70,33%	72,63%	67,86%	72,31%	70,17%	68,35%
Classifier	LDA	61,78%	63,53%	54,23%	62,14%	59,89%	57,50%
	SVM	66,58%	70,47%	66,48%	68,45%	67,57%	67,51%
	MLP	68,71%	70,87%	65,61%	69,21%	67,96%	68,01%
Feature Set	MAG	71,17%	73,77%	66,26%	71,75%	70,17%	69,28%
	PHA	62,54%	64,05%	56,04%	61,34%	60,49%	60,79%
	M&P	69,70%	72,74%	67,29%	71,25%	69,53%	68,94%
	IMP	59,37%	62,59%	58,83%	62,06%	60,38%	58,34%
		Antenna	MD	67,01%	66,93%	63,13%	
			MH	69,35%	65,59%	69,92%	
			AV	63,45%	62,90%	59,96%	

Figure 7: Contingency tables for the classification recognition rates. Abbreviations used for features sets: MAG for spectral magnitude, PHA for phase, M&P for concatenation of spectral magnitude and phase, and IMP for impulse response. The color code goes from white (lowest value) to red (highest value) for each pair of variables.

different effects on different antennas. As random effects, we considered intercepts for speaker (with the session nested under it), classifier and feature set, as well as random slopes for the effect of antenna type on these variables. We inspected the residuals of the model and did not notice any sign of deviation from homoscedasticity or normality.

To obtain p-values, we performed likelihood ratio tests between the full model described above and models without individual fixed effects of antenna type and position, but with the same random effects. Antenna type ($\chi^2(6) = 41.056$, $p = 2.823 \times 10^{-7}$) and position ($\chi^2(6) = 49.668$, $p = 5.479 \times 10^{-9}$) had significant effects on the phoneme recognition rate. Increases of $2.7\% \pm 2.8\%$ and $6.2\% \pm 3.0\%$ were estimated by the model when switching from the AV antenna to the MD and the MH antenna, respectively. Likewise for the antenna positions, decreases of $0.1\% \pm 0.9\%$ and $2.8\% \pm 0.9\%$ were estimated when changing from position A to positions B and C, respectively.

Before discussing the results, it is important to highlight the limitations of this study. The sample size of investigated antenna types and positions did not exhaust the set of possibilities, as new antennas may be designed in the future and more positions may be introduced into our system, e.g., to capture nasal coupling and/or glottal movement. Additionally, the antenna types AV and MH have SMA connectors (SubMiniature version A) which hinders the contact of their full surface with the skin (see Figure 1), creating an air gap between the antenna and the skin, changing its radiation characteristics [25]. Since the MD antenna's JSC connector does not introduce this effect, this may have influenced the results in this study. The JSC connector, however, introduces the need for an SMA-JSC adapter, since the SSI's hardware is equipped with SMA connectors, and also for cables connecting the adapter to the antenna (see top row of Figure 2). In comparison with the SMA-cables (seen in Figure 3), these JSC-cables are more prone to breakage and less shielded. This results in a stronger coupling between JSC-cables, which introduces noise into the measurement. This effect should be investigated further and, if possible, minimized in next studies with our SSI. Besides that, this study was more concerned with comparing antenna types and positions than with achieving the highest possible recognition rate. Therefore, rather simple types of classifiers were used, and the hyper-parameter optimization was not exhaustively performed.

From this study can be concluded that our radar-based SSI achieved better recognition rates with monopole antennas than with Vivaldi antennas. This may be due to differences in radiation characteristic. Monopole antennas radiate in the direction of the vocal tract, whereas Vivaldi antennas also radiate tangentially to the skin. Additionally, this improvement is more notable with the MH antenna than with the MD. Analysing the t-values (ratio between estimated slope and its standard error) estimated by the LMEM, MD antenna had $t = 0.972$ and MH, $t = 2.037$. The low value for the MD antenna may be due to the higher variance of the AV antenna, despite its lower mean accuracy for all positions.

Different antenna positions also yielded different recognition rates, depending on the antenna type. Positions A and B were superior compared to position C, when using antennas AV and MD. However, the trend observed with antenna MH was different: positions A and C achieved the best results over position B. This suggests that our SSI's performance may vary in a non-linear manner with the distance between the antenna and the lips, and could explain the low t-value estimated for position B ($t = -0.129$) in comparison with position C ($t = -3.008$). An analysis on how different phonemes were recognised with the different antenna positions may shed more light into the issue, as different phonemes have different places of articulation and some might have been better accounted for depending on the position of the antenna.

5. Acknowledgment

This research work is funded by the European Regional Development Fund (EFRE, Project "Radar Speech"), by the Sächsische Aufbaubank (SAB) under grant number 100328626 and by the German Federal Ministry of Education and Research (BMBF) under grant number 20D1930B.

6. References

- [1] G. S. Meltzner, J. T. Heaton, Y. Deng, G. De Luca, S. H. Roy, and J. C. Kline, "Development of sEMG sensors and algorithms for

- silent speech recognition,” *Journal of neural engineering*, vol. 15, no. 4, 2018.
- [2] Y. Wang, M. Zhang, R. Wu, H. Gao, M. Yang, Z. Luo, and G. Li, “Silent speech decoding using spectrogram features based on neuromuscular activities,” *Brain Sciences*, vol. 10, no. 7, 2020. [Online]. Available: <https://www.mdpi.com/2076-3425/10/7/442>
 - [3] M. Angrick, C. Herff, E. Mugler, M. C. Tate, M. W. Slutzky, D. J. Krusienski, and T. Schultz, “Speech synthesis from ECoG using densely connected 3d convolutional neural networks,” *Journal of Neural Engineering*, vol. 16, no. 3, 2019.
 - [4] G. K. Anumanchipalli, J. Chartier, and E. F. Chang, “Speech synthesis from neural decoding of spoken sentences,” *Nature*, vol. 568, pp. 493–498, 2019.
 - [5] J. A. Gonzalez, L. A. Cheah, A. M. Gomez, P. D. Green, J. M. Gilbert, S. R. Ell, R. K. Moore, and E. Holdsworth, “Direct speech reconstruction from articulatory sensor data by machine learning,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 12, pp. 2362–2374, 2017.
 - [6] S. Stone and P. Birkholz, “Cross-speaker silent-speech command word recognition using electro-optical stomatography,” in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 7849–7853.
 - [7] B. Denby, T. Schultz, K. Honda, T. Hueber, J. Gilbert, and J. Brumberg, “Silent speech interfaces,” *Speech Communication*, vol. 52, no. 4, pp. 270–287, 2010.
 - [8] T. Schultz, M. Wand, T. Hueber, D. J. Krusienski, C. Herff, and J. S. Brumberg, “Biosignal-based spoken communication: A survey,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 12, pp. 2257–2271, 2017.
 - [9] J. A. Gonzalez-Lopez, A. Gomez-Alanis, J. M. Martín Doñas, J. L. Pérez-Córdoba, and A. M. Gomez, “Silent speech interfaces for speech restoration: A review,” *IEEE Access*, vol. 8, pp. 177 995–178 021, 2020.
 - [10] J. M. Gilbert, J. A. Gonzalez, L. A. Cheah, S. R. Ell, P. Green, R. K. Moore, and E. Holdsworth, “Restoring speech following total removal of the larynx by a learned transformation from sensor data to acoustics,” *The Journal of the Acoustical Society of America*, vol. 141, no. 3, pp. EL307–EL313, 2017. [Online]. Available: <https://doi.org/10.1121/1.4978364>
 - [11] J. F. Holzrichter, G. C. Burnett, and L. C. Ng, “Speech articulator measurements using low power EM-wave sensors,” *The Journal of the Acoustical Society of America*, vol. 103, no. 1, pp. 622–625, 1998.
 - [12] L. Wen, C. Gu, and J.-F. Mao, “Silent speech recognition based on short-range millimeter-wave sensing,” in *2020 IEEE/MTT-S International Microwave Symposium (IMS)*, 2020, pp. 779–782.
 - [13] D. Ferreira, S. Silva, F. Curado, and A. Teixeira, “Exploring silent speech interfaces based on frequency-modulated continuous-wave radar,” *Sensors*, vol. 22, no. 2, 2022. [Online]. Available: <https://www.mdpi.com/1424-8220/22/2/649>
 - [14] P. Birkholz, S. Stone, K. Wolf, and D. Plettemeier, “Non-invasive silent phoneme recognition using microwave signals,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 12, pp. 2404–2411, 2018.
 - [15] P. Amini Digebsara, C. Wagner, P. Schaffer, M. Bärhold, S. Stone, D. Plettemeier, and P. Birkholz, “On the optimal set of features and the robustness of classifiers in radar-based silent phoneme recognition,” in *Studenttexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2021*, S. Hillmann, B. Weiss, T. Michael, and S. Möller, Eds. TUDpress, Dresden, 2021, pp. 112–119.
 - [16] C. Wagner, P. Schaffer, P. Amini Digebsara, M. Bärhold, D. Plettemeier, and P. Birkholz, “Silent speech command word recognition using stepped frequency continuous wave radar,” *Scientific Reports*, vol. 12, no. 4192, 2022.
 - [17] X. Fang, M. Ramzan, Q. Wang, and D. Plettemeier, “Compact antipodal vivaldi antennas for body area communication,” in *Advances in Body Area Networks I*. Springer, 2019, pp. 357–369.
 - [18] X. Fang, M. Ramzan, Q. Zhang, S. Pérez-Simbor, Q. Wang, N. Neumann, C. García-Pardo, N. Cardona, and D. Plettemeier, “Experimental in-body to on-body and in-body to in-body path loss models of planar elliptical ring implanted antenna in the ultra-wide band,” in *2019 13th International Symposium on Medical Information and Communication Technology (ISMICT)*, 2019, pp. 1–5.
 - [19] P. S. Hall, Y. Hao, Y. I. Nechayev, A. Alomainy, C. C. Constantinou, C. Parini, M. R. Kamarudin, T. Z. Salim, D. T. Hee, R. Dubrovka, A. S. Owadally, W. Song, A. Serra, P. Nepa, M. Gallo, and M. Bozzetti, “Antennas and propagation for on-body communication systems,” *IEEE Antennas and Propagation Magazine*, vol. 49, no. 3, pp. 41–58, 2007.
 - [20] L. Berkelmann and D. Manteuffel, “Antenna parameters for on-body communications with wearable and implantable antennas,” *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 9, pp. 5377–5387, 2021.
 - [21] P. Boersma and D. Weenink, “Praat: doing phonetics by computer [computer program], version 6.2.09, retrieved 15 february 2022 from <http://www.praat.org/>,” 2022.
 - [22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
 - [23] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022. [Online]. Available: <http://www.R-project.org/>
 - [24] D. Bates, M. Mächler, B. Bolker, and S. Walker, “Fitting linear mixed-effects models using lme4,” *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.
 - [25] K. Sasaki, M. Mizuno, K. Wake, and S. Watanabe, “Measurement of the dielectric properties of the skin at frequencies from 0.5 GHz to 1 THz using several measurement systems,” in *2015 40th International Conference on Infrared, Millimeter, and Terahertz waves (IRMMW-THz)*, 2015, pp. 1–2.