Análisis de Señales para la Detección de Patologías en Voz: Proyecto Cordectomía

Alfonso Gamboa Rubén,
Flores Monteros Edsel Yetlanezi

Resumen—La cordectomía, procedimiento quirúrgico que implica la extirpación parcial o total de los pliegues vocales, compromete severamente la capacidad comunicativa del paciente, afectando su identidad y calidad de vida. Este proyecto presenta el diseño y evaluación de un sistema de procesamiento digital de señales (DSP) orientado a la rehabilitación vocal no invasiva mediante la reconstrucción espectral. La metodología evolucionó a través de tres fases iterativas: una aproximación inicial en el dominio de la frecuencia (FFT global), un modelo adaptativo basado en metadatos y filtrado de intensidad adaptable, y finalmente, la implementación basada en la Transformada de Fourier de Tiempo Corto (STFT) y estimadores estadísticos (MMSE-STSA). Los resultados experimentales demostraron que, si bien la sustracción de ruido estacionario mediante algoritmos de Wiener y Ephraim-Malah es efectiva, la reconstrucción de la voz requiere una intervención más compleja a nivel de las micro-características que forman la voz para lograr preservar la identidad del paciente y evitar artefactos o distorsiones. El estudio concluye proponiendo una versión adicional de experimentación modular que implementa herramientas de inteligencia artificial.

Index Terms—Procesamiento Digital de Señales (DSP), Transformada de Fourier de Tiempo Corto (STFT), Filtro de Wiener, Filtro Savitzky-Golay, Detección de Actividad de Voz (VAD), Análisis Espectral, Rehabilitación Fónica, Python, Cordectomía, Ephraim-Malah, Formantes, Inteligencia Artificial (IA), RLHF, Red Neuronal, Speech Emotion Recognition (SER).

Objetivo General

Desarrollar y evaluar algoritmos de procesamiento digital de señales basado en análisis espectral de tiempo corto y modelado estadístico, en relación a la capacidad de mejorar la calidad de la voz y restaurar parcialmente las características tímbricas en grabaciones de voz de pacientes sometidos a cordectomía.

Objetivos Específicos

1. **Caracterización Acústica:** Construir una base de datos pareada (pre y post-operatoria) para identificar los patrones de pérdida armónica y deformación espectral en el dominio de la frecuencia causados por la intervención quirúrgica.
2. **Optimización de la Relación Señal-Ruido (SNR):** Implementar y comparar técnicas de sustracción espectral (Noisereduce vs. Ephraim-Malah/VAD) para minimizar el ruido estacionario inherente a la fonación soplada sin degradar los transitorios de la voz.
3. **Reconstrucción Espectral:** Experimentar con algoritmos de transferencia de características que utilicen una máscara espectral diferencial (T_{dB}) para proyectar el timbre e identidad del sonido vocal (envolvente de frecuencia de la voz) sano sobre la señal patológica.
4. **Validación Técnica:** Evaluar mediante espectrogramas y gráficas comparativas, la efectividad de los algoritmos en la rehabilitación de formantes y reducción de artefactos y desfase de frecuencias armónicas.

Documento entregado el dia: 10 de diciembre de 2025. Trabajo realizado para la materia de Análisis de Señales, Universidad Veracruzana.

Abstract—Cordectomy, a surgical procedure involving partial or total removal of vocal folds, severely compromises communicative capacity and patient identity. This project presents the design and evaluation of a digital signal processing (DSP) system for non-invasive vocal rehabilitation via spectral reconstruction. The methodology evolved through three iterative phases: an initial frequency domain approach (global FFT), an adaptive model based on metadata, and finally, an implementation based on Short-Time Fourier Transform (STFT) with statistical estimators (MMSE-STSA). Experimental results showed that while stationary noise subtraction via Wiener and Ephraim-Malah algorithms is effective, voice reconstruction requires complex intervention at the micro-feature level to preserve patient identity and avoid artifacts. The study concludes by proposing a future modular version implementing artificial intelligence tools.

Index Terms—Procesamiento Digital de Señales (DSP), Transformada de Fourier de Tiempo Corto (STFT), Filtro de Wiener, Filtro Savitzky-Golay, Detección de Actividad de Voz (VAD), Análisis Espectral, Rehabilitación Fónica, Python, Cordectomía, Ephraim-Malah, Formantes, Inteligencia Artificial (IA), RLHF, Red Neuronal, Speech Emotion Recognition (SER).

1. Introducción

LAlo

2. Metodología

Referencias

[1] T. E. Oliphant, .A guide to NumPy,ÜSA: Trelgol Publishing, vol. 1, 2006.
[2] W. McKinney, "Python for data analysis: Data wrangling with Pandas, NumPy, and IPython,.O'Reilly Media, Inc., 2012.
[3] S. van der Walt, S. C. Colbert, and G. Varoquaux, "The NumPy array: A structure for efficient numerical computation,"*Computing in Science & Engineering*, vol. 13, no. 2, pp. 22-30, 2011.
[4] J. M. Kizza, "Python for scientific computing,ïn *Guide to Computer Network Security*, Springer, 2017, pp. 263-283.
[5] P. Virtanen et al., "SciPy 1.0: fundamental algorithms for scientific computing in Python,"*Nature Methods*, vol. 17, no. 3, pp. 261-272, 2020.
[6] E. Jones, T. Oliphant, and P. Peterson, "SciPy: Open source scientific tools for Python,"2001.
[7] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures,"*Analytical Chemistry*, vol. 36, no. 8, pp. 1627-1639, 1964.
[8] R. W. Schafer, "What is a Savitzky-Golay filter? [lecture notes],"*IEEE Signal Processing Magazine*, vol. 28, no. 4, pp. 111-117, 2011.
[9] W. H. Press and S. A. Teukolsky, "Savitzky-Golay smoothing filters,"*Computers in Physics*, vol. 4, no. 6, pp. 669-672, 1990.
[10] M. Schmid, D. Rath, and U. Diebold, "Why and how Savitzky–Golay filters should be replaced,"*ACS Measurement Science Au*, vol. 2, no. 2, pp. 185-196, 2022.
[11] H. H. Madden, Çomments on the Savitzky-Golay convolution method for least-squares-fit smoothing and differentiation of digital data,"*Analytical Chemistry*, vol. 50, no. 9, pp. 1383-1386, 1978.
[12] J. O. Smith, "Spectral audio signal processing,"W3K Publishing, 2011.
[13] L. R. Rabiner and B. Gold, "Theory and application of digital signal processing,.Englewood Cliffs, NJ: Prentice-Hall, Inc., 1975.
[14] J. B. Allen and L. R. Rabiner, .A unified approach to short-time Fourier analysis and synthesis,"*Proceedings of the IEEE*, vol. 65, no. 11, pp. 1558-1564, 1977.
[15] M. R. Portnoff, "Time-frequency representation of digital signals and systems based on short-time Fourier analysis,"*IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 1, pp. 55-69, 1980.
[16] M. Dolson, "The phase vocoder: A tutorial,"*Computer Music Journal*, vol. 10, no. 4, pp. 14-27, 1986.
[17] D. Griffin and J. Lim, "Signal estimation from modified short-time Fourier transform,"*IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 2, pp. 236-243, 1984.
[18] B. Sharpe, Ïnvertibility of overlap-add processing,"https://gauss256.github.io/blog/cola.html, accessed July 2019.
[19] L. R. Rabiner and R. W. Schafer, "Digital processing of speech signals,.Englewood Cliffs, NJ: Prentice Hall, 1978.
[20] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator,"*IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1109-1121, 1984.
[21] N. Wiener, .Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications,"MIT Press, 1949.

[22] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter,"*IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1218-1234, 2006.

[23] P. C. Loizou, "Speech enhancement: theory and practice,ÇRC Press, 2013.

[24] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction,"*IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 2, pp. 113-120, 1979.

[25] J. Sohn, N. S. Kim, and W. Sung, .A statistical model-based voice activity detection,"*IEEE Signal Processing Letters*, vol. 6, no. 1, pp. 1-3, 1999.

[26] A. J. M. Houtsma, "Pitch and timbre: Definition, meaning and use,"*Journal of New Music Research*, vol. 26, no. 2, pp. 104-115, 1997.

[27] H. M. Teager and S. M. Teager, .Evidence for nonlinear sound production mechanisms in the vocal tract,ïn *Speech Production and Speech Modelling*, Springer, 1990, pp. 241-261.

[28] P. Ladefoged, "Vowels and consonants: An introduction to the sounds of languages,"Malden, MA: Blackwell Publishers, 2001.

[29] G. Fant, .Acoustic theory of speech production,"The Hague: Mouton, 1960.

[30] G. E. Peterson and H. L. Barney, Çontrol methods used in a study of the vowels,"*The Journal of the Acoustical Society of America*, vol. 24, no. 2, pp. 175-184, 1952.

[31] J. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler, .Acoustic characteristics of American English vowels,"*The Journal of the Acoustical Society of America*, vol. 97, no. 5, pp. 3099-3111, 1995.

[32] K. N. Stevens, .Acoustic phonetics,"MIT Press, 1998.

[33] D. H. Whalen and A. G. Levitt, "The universality of intrinsic F0 of vowels,"*Journal of Phonetics*, vol. 23, no. 3, pp. 349-366, 1995.

[34] I. R. Titze, "Principles of voice production,Ïowa City: National Center for Voice and Speech, 2000.

[35] M. Hirano, Çlinical examination of voice,"Springer Science & Business Media, 2013.

[36] C. E. Silver et al., Çurrent trends in initial management of laryngeal cancer,"*European Archives of Oto-Rhino-Laryngology*, vol. 266, no. 9, pp. 1333-1352, 2009.

[37] M. Remacle et al., .Endoscopic cordectomy. A proposal for a classification,"*European Archives of Oto-Rhino-Laryngology*, vol. 257, no. 4, pp. 227-231, 2000.

[38] E. V. Sjögren et al., "Voice outcome in T1a midcord glottic carcinoma,"*Archives of Otolaryngology–Head & Neck Surgery*, vol. 134, no. 9, pp. 965-972, 2008.

[39] T. Yılmaz et al., "Voice after cordectomy type I or type II or radiation therapy,"*Otolaryngology–Head and Neck Surgery*, vol. 168, no. 3, pp. 559-568, 2023.

[40] L. M. Aaltonen et al., "Voice quality after treatment of early vocal cord cancer,"*International Journal of Radiation Oncology Biology Physics*, vol. 90, no. 2, pp. 255-270, 2014.

[41] H. S. Lee et al., "Voice outcome according to surgical extent of transoral laser microsurgery,"*The Laryngoscope*, vol. 126, no. 9, pp. 2051-2056, 2016.

[42] A. K. Fouad et al., "Laryngeal compensation for voice production after CO2 laser cordectomy,"*Clinical and Experimental Otorhinolaryngology*, vol. 8, no. 4, pp. 340-346, 2015.

[43] G. Fant, .Acoustic theory of speech production: with calculations based on X-ray studies,"The Hague: Mouton, 1960.

[44] T. Chiba and M. Kajiyama, "The vowel: Its nature and structure,"Tokyo-Kaiseikan Publishing Co., 1941.

[45] K. N. Stevens, .Acoustic phonetics,Çurrent Studies in Linguistics Series, vol. 30, MIT Press, 1999.

[46] J. L. Flanagan, "Speech analysis synthesis and perception,"Berlin: Springer-Verlag, 1972.

[47] I. R. Titze, "Nonlinear source-filter coupling in phonation: Theory,"*The Journal of the Acoustical Society of America*, vol. 123, no. 5, pp. 2733-2749, 2008.

[48] P. Birkholz, D. Jackèl, and B. J. Kröger, Çonstruction and control of a three-dimensional vocal tract model,ïn *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, IEEE, vol. 1, 2006.

[49] B. H. Story, .A parametric model of the vocal tract area function,"*The Journal of the Acoustical Society of America*, vol. 117, no. 5, pp. 3231-3254, 2005.

[50] W. J. Hardcastle, J. Laver, and F. E. Gibbon, *The Handbook of Phonetic Sciences*, 2nd ed. Oxford: Wiley-Blackwell, 2010.

[51] I. Goodfellow, Y. Bengio, y A. Courville, *Deep Learning*. MIT Press, 2016.

[52] J. Wang, K. Chin, y H. Wang, "Speaker-informed speech enhancement and separation,.en *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2021.

[53] Y. Fathullah et al., "Neural Speech Synthesis using Semantic Tokens,"*arXiv preprint arXiv:2305.xxxx*, 2023.

[54] W.-N. Hsu et al., "HuBERT: Self-Supervised Speech Representation Learning,.en *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 3451-3460, 2021.

[55] K. Qian et al., ÇontentVec: An Improved Self-Supervised Speech Representation,.en *Proc. of the 39th International Conference on Machine Learning (ICML)*, 2022.

[56] N. Tishby y N. Zaslavsky, "Deep learning and the information bottleneck principle,.en *IEEE Information Theory Workshop (ITW)*, 2015.

[57] X. Tan et al., .A Survey on Neural Speech Synthesis,"*arXiv preprint arXiv:2106.15561*, 2021.

[58] RVC-Project, Retrieval-based Voice Conversion WebUI,"GitHub repository, 2023.

[59] C. Kavin (svc-develop-team), "So-VITS-SVC: SoftVC VITS Singing Voice Conversion,"GitHub repository, 2023.

[60] E. Gölge et al., Çoqui XTTS: Open-Source Text-to-Speech Model,Çoqui AI, 2023.

[61] A. Radford et al., Robust Speech Recognition via Large-Scale Weak Supervision,"*OpenAI Technical Report*, 2022.

[62] J. Kong, J. Kim, y J. Bae, "HiFi-GAN: Generative Adversarial Networks for Efficient and High Fidelity Speech Synthesis,.en *Proc. NeurIPS*, 2020.

[63] P. Christiano et al., "Deep Reinforcement Learning from Human Feedback,"*Advances in Neural Information Processing Systems*, 2017.

[64] R. A. Khalil et al., "Speech Emotion Recognition Using Deep Learning Techniques: A Review,"*IEEE Access*, vol. 7, pp. 117327-117345, 2019.