



## Análise da relação entre a insegurança e a variação de intensidade da voz

Aluno: Luiz Henrique Costa Freitas, l202403@dac.unicamp.br

Orientador: Prof. Dr. Tiago Fernandes Tavares

Coorientadora: Profa. Dra. Paula D. Paro Costa

### I. INTRODUÇÃO

Métodos computacionais para a análise afetiva da voz têm recebido grande foco científico nos últimos anos. Tais métodos se relacionam a diversas áreas, como linguística [1], ciência forense [2], ciências médicas [3] [4] dentre outras [5]. Neste trabalho, analisamos a relação entre a variação da intensidade da voz e o sentimento de medo e insegurança. Esta análise teve foco em avaliar uma possível aplicação no *feedback* automático para treinamento de palestrantes.

A voz possui diversas características acústicas, cada uma relacionada a algum aspecto específico. Direcionamos nosso estudo para a variação da intensidade da voz, que é medida através do shimmer e é considerada uma característica importante na classificação automática de emoções na voz [5]. Durante a pesquisa, buscamos entender qual a percepção psico-acústica do shimmer e se ele está de alguma maneira relacionado com a voz trêmula e sentimento de insegurança.

Utilizamos inicialmente uma base de dados composta por arquivos de áudio gravados por atores representando diferentes sentimentos[6]. Após, criamos uma segunda base de dados com falas gravadas por não-atores, usando vídeos disponíveis no Youtube. Extraímos o shimmer para cada arquivo de áudio presente nas bases de dados e analisamos os resultados obtidos.

Verificamos através dos valores obtidos para a primeira base de dados que a média do shimmer no sentimento de 'medo' foi maior do que em comparação com os demais. Contudo, para a segunda base de dados não obtivemos uma separação perceptível. Fizemos, então, gravações em estúdio para investigar a percepção do shimmer e encontramos dificuldades com relação ao significado do shimmer na voz. Isso nos mostrou que o shimmer não possui uma relação clara com a insegurança na voz.

### II. MÉTODO E RESULTADOS

Iniciamos a pesquisa analisando a primeira base de dados[6]. Ela é composta por arquivos de áudio em que atores representam seis sentimentos: raiva, nojo, medo, alegria, tristeza e surpresa. A base possui 6 gravações representando

cada sentimento, exceto para o sentimento 'surpresa', que possui apenas 5 arquivos. A Figura 1 mostra as distribuições dos valores de shimmer estimados em cada uma das categorias de emoção. Podemos observar que o shimmer para a emoção "medo" é tipicamente mais alto que para as outras emoções. Isso indica que os atores buscaram expressar medo através da tremulação da voz.

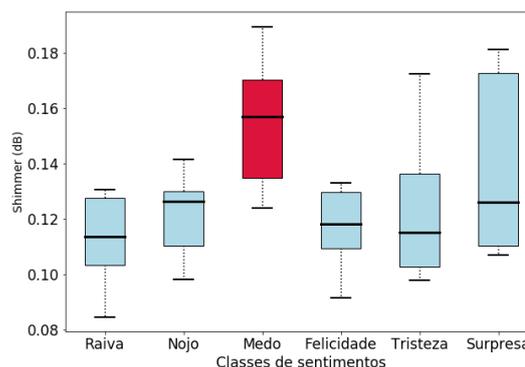


Fig. 1. BoxPlot mostrando a distribuição estatística do shimmer para as emoções da base de dados.

Com o intuito de verificar a abrangência desse resultado em outras circunstâncias, criamos uma base de dados a partir de arquivos de áudio disponíveis na internet em que os locutores estão em diferentes situações. Escolhemos locutores com experiência na oratória, jornalistas e youtubers, e outros menos experientes, alunos apresentando trabalho de conclusão de curso (TCC), de forma que houvesse na base de dados arquivos representando pessoas confiantes e inseguras na sua fala. O resultado obtido para essa análise é mostrado na Figura 2 e revela que nesse caso não houve uma distribuição que permitisse uma diferenciação entre as classes através do shimmer.

Os arquivos obtidos apresentaram uma grande variação na qualidade do áudio, o que pode ter sido um dos motivos

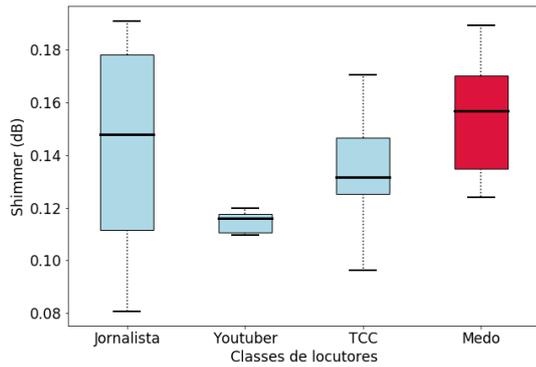


Fig. 2. BoxPlot do shimmer extraído para cada classe da base de dados.

dos resultados não serem condizentes com a análise feita anteriormente. Alguns deles, por exemplo, possuíam um ruído de fundo considerável e que pode ter afetado no cálculo do shimmer. Tendo em vista os resultados apresentados, utilizamos um estúdio para gravação de maneira controlada com o intuito de verificar a nossa percepção psico-acústica do shimmer na voz seguindo o processo mostrado na Figura 3. Durante as gravações, gravamos os membros desta pesquisa tentando, intencionalmente, fazer a voz trêmula. Porém, não temos medições objetivas do trêmulo. Medimos, então, o shimmer de cada gravação, e comparamos os valores com aqueles relacionados às bases anteriores.



Fig. 3. Fluxograma mostrando o processo de gravação até análise.

Os arquivos com a voz mais trêmula apresentaram valores menores de shimmer em comparação com os demais, conforme mostra a Figura 4, indicando que a percepção do shimmer não é tão intuitiva. Contrastamos nossos resultados com pesquisas que utilizam o shimmer para identificar atitudes de gentileza e hostilidade na fala [1] e gêneros [7]. Ambas utilizaram o shimmer, dentre outras *features*, porém não obtiveram valores significativos para classificação através do shimmer.

### III. CONCLUSÕES

Os resultados apresentados indicam que o shimmer não possui sentido prático dentro da análise de insegurança na voz durante a fala corrida, já que deve ser extraído para vogais sustentadas [8]. Além disso, também percebemos que a percepção psico-acústica do shimmer não está relacionada com a voz trêmula. Considerando as possíveis aplicações da

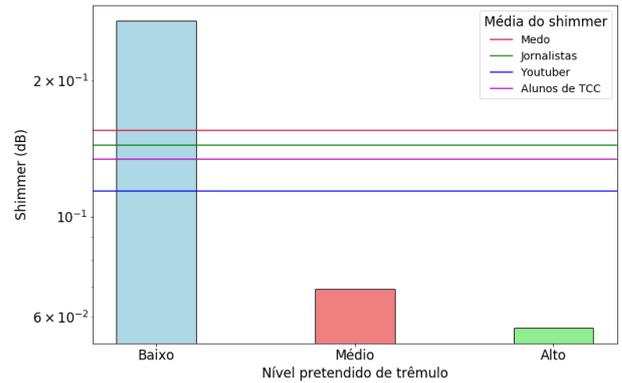


Fig. 4. Relação entre o Shimmer e o trêmulo pretendido da voz

identificação de insegurança na voz, entendemos, portanto, que a utilização do shimmer não possui grande relevância.

### REFERENCES

- [1] P. A. SIQUEIRA, Júlia e BARBOSA, “Diferenças prosódicas em atos combinados a atitudes distintas,” *Anais do I Congresso Brasileiro de Prosódia*.
- [2] A. Ferreira and V. Fernandes, “Consistency of the F0, Jitter, Shimmer and HNR voice parameters in GSM and VOIP communication,” in *2017 22nd International Conference on Digital Signal Processing (DSP)*, (London), pp. 1–5, IEEE, Aug. 2017.
- [3] J. P. Teixeira and A. Gonçalves, “Algorithm for Jitter and Shimmer Measurement in Pathologic Voices,” *Procedia Computer Science*, vol. 100, pp. 271–279, 2016.
- [4] M. Brockmann, M. J. Drinnan, C. Storck, and P. N. Carding, “Reliable Jitter and Shimmer Measurements in Voice Clinics: The Relevance of Vowel, Gender, Vocal Intensity, and Fundamental Frequency Effects in a Typical Clinical Task,” *Journal of Voice*, vol. 25, pp. 44–53, Jan. 2011.
- [5] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. Andre, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan, and K. P. Truong, “The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing,” *IEEE Transactions on Affective Computing*, vol. 7, pp. 190–202, Apr. 2016.
- [6] P. D. P. Costa, *Two-dimensional expressive speech animation*. PhD thesis, School of Electrical and Computer Engineering, 2015.
- [7] J. P. Teixeira and P. O. Fernandes, “Jitter, Shimmer and HNR Classification within Gender, Tones and Vowels in Healthy Voices,” *Procedia Technology*, vol. 16, pp. 1228–1237, 2014.
- [8] J. M. d. S. Lopes, “Ambiente de análise robusta dos principais parâmetros qualitativos da voz,” 2009.