



RESEARCH STAY WEEK 9,

**► Attention in
Sequence-to-Sequence
Models**

Daniel Cajas A01708637

► CONTEXT

A concept used for enhancing seq2seq models is attention. It provides a new path for context to get to the decoder, allowing a better identification of long distance relationships. It is so powerful that it inspired the transformer architecture (the T in GPT), which throws away the RNNs normally used to get the sequence information and uses only self attention to encode context.

► SEARCH METHODOLOGY

I am very interested in transformers but advanced transformer architectures are above the scope of this week's work. That is why i decided to focus on speech emotion recognition (SER). As [1] mentions, attention in SER does show some improvements but the results were quite inconsistent. Having seen the potential of attention in other ML tasks i found it weird that it did not help SER. I wanted to find newer research that implemented attention in a clear and beneficial manner.

► COMPARISON

Title	Objective	Year	Input format	Architecture	Takeaways
A Review on Speech Emotion Recognition Using Deep Learning and Attention Mechanism	Sumarizing state of the art SER models and comparing them. Noting the use of attention and how it affected performance.	2021	Mostly spectrogram based with some using some custom features extracted by an embedding preprocessing step.	A variety of CNNs, RNNs, regrssion models and SVMs	While attention does show some potential, its immediate benefits are not quite as clear as in other fields
Improved ShuffleNet V2 network with attention for speech emotion recognition	Upgrtading the ShuffleNet V2 network using attention modules and a new activation function to increase performance while keeping computational costs relativley low.	2024	Histogram	Shufflenet v2 with hardwish and efficient channel attention (ECA) module.	Shows significant improvements (3% accuracy improvement) when integrating atenttion into well known arqitutures.

► BIBLIOGRAFÍA

[1] Lieskovská, E., Jakubec, M., Jarina, R., & Chmuliík, M. (2021). A review on speech emotion recognition using deep learning and attention mechanism [Review]. Electronics (Switzerland), 10(10). <https://doi.org/10.3390/electronics10101163>

[2] Udeh, C. P., Chen, L., Du, S., Liu, Y., Li, M., & Wu, M. (2025). Improved ShuffleNet V2 network with attention for speech emotion recognition [Article]. Information Sciences, 689. <https://doi.org/10.1016/j.ins.2024.121488>