		doc_1	doc_2		decision	id
cases	authors	- Jui Shah - Yaman Kumar Singla - Changyou Chen - Rajiv Ratn Shah	authors	- Jui Shah - Yaman Kumar Singla - Changyou Chen R. Shah	DUPLICATES	232
	title	What all do audio transformer models hear? Probing Acoustic Representations for Language Delivery and its Structure	title	What all do audio transformer models hear? Probing Acoustic Representations for Language Delivery and its Structure		
	publication_date	2021-01-02 06:29:12+00:00	publication_date	2021-01-02 00:00:00		
	source	SupportedSources.ARXIV	source	SupportedSources.SEMANTIC_SCHOLAR		
	journal	None	journal	ArXiv		
	volume		volume	abs/2101.00387		
	doi		doi	10.20944/preprints202101.0081.v1		
	urls	- http://arxiv.org/pdf/2101.00387v2 - http://arxiv.org/abs/2101.00387v2 - http://arxiv.org/pdf/2101.00387v2	urls	https://www.semanticscholar.org/paper/5e32055b61bf58c336af41ca8642c7138a753aa7		
	id	id-7345381725309213464	id	id3287545049555777753		
	abstract	In recent times, BERT based transformer models have become an inseparable part of the 'tech stack' of text processing models. Similar progress is being observed in the speech domain with a multitude of models observing state-of-the-art results by using audio transformer models to encode speech. This begs the question of what are these audio transformer models learning. Moreover, although the standard methodology is to choose the last layer embedding for any downstream task, but is it the optimal choice? We try to answer these questions for the two recent audio transformer models, Mockingjay and wave2vec2.0. We compare them on a comprehensive set of language delivery and structure features including audio, fluency and pronunciation features. Additionally, we probe the audio models' understanding of textual surface, syntax, and semantic features and compare them to BERT. We do this over exhaustive settings for native, non-native, synthetic, read and spontaneous speech datasets	abstract	In recent times, BERT based transformer models have become an inseparable part of the â€˜tech stackâ€™ of text processing models. Similar progress is being observed in the speech domain with a multitude of models observing state-of-the-art results by using audio transformer models to encode speech. This begs the question of what are these audio transformer models learning. Moreover, although the standard methodology is to choose the last layer embeddings for any downstream task, but is it the optimal choice? We try to answer these questions for the two recent audio transformer models, Mockingjay and wave2vec2.0 . We compare them on a comprehensive set of language delivery and structure features including audio, fluency and pronunciation features. Additionally, we probe the audio modelsâ€™ understanding of textual surface, syntax, and semantic features and compare them to BERT. We do this over exhaustive settings for native, nonnative, synthetic, read and spontaneous speech datasets.		
	versions		versions			