

cases	doc_1		doc_2		decision	id
	authors	<ul style="list-style-type: none"><li>Baevski, Alexei</li><li>De Seyssel, Maureen</li><li>Dunbar, Ewan</li><li>Dupoux, Emmanuel</li><li>Kharitonov, Evgeny</li><li>Nguyen, Tu Anh</li><li>Rivi�re, Morgane</li><li>Roz�, Patricia</li></ul>	authors	<ul style="list-style-type: none"><li>Tu Anh Nguyen</li><li>Maureen de Seyssel</li><li>Patricia Roz�</li><li>Morgane Rivi�re</li><li>Evgeny Kharitonov</li><li>Alexei Baevski</li><li>Ewan Dunbar</li><li>Emmanuel Dupoux</li></ul>	DUPLICATES	293
	title	The Zero Resource Speech Benchmark 2021: Metrics and baselines for unsupervised spoken language modeling	title	The Zero Resource Speech Benchmark 2021: Metrics and baselines for unsupervised spoken language modeling		
	publication_date	2020-12-06 00:00:00	publication_date	2020-12-01 00:00:00		
	source	SupportedSources.CORE	source	SupportedSources.INTERNET_ARCHIVE		
	journal		journal			
	volume		volume			
	doi	None	doi			
	urls	<ul style="list-style-type: none"><li>https://core.ac.uk/download/362228848.pdf</li></ul>	urls	<ul style="list-style-type: none"><li>https://web.archive.org/web/20201126222115/https://arxiv.org/pdf/2011.11588v1.pdf</li></ul>		
	id	id-7233653472157712228	id	id1450823174297676281		
	abstract	14 pages, including references and supplementary materialInternational audienceWe introduce a new unsupervised task, spoken language modeling: the learning of linguistic representations from raw audio signals without any labels, along with the Zero Resource Speech Benchmark 2021: a suite of 4 black-box, zero-shot metrics probing for the quality of the learned models at 4 linguistic levels: phonetics, lexicon, syntax and semantics. We present the results and analyses of a composite baseline made of the concatenation of three unsupervised systems: self-supervised contrastive representation learning (CPC), clustering (k-means) and language modeling (LSTM or BERT). The language models learn on the basis of the pseudo-text derived from clustering the learned representations. This simple pipeline shows better than chance performance on all four metrics, demonstrating the feasibility of spoken language modeling from raw speech. It also yields worse performance compared to text-based 'topline' systems trained on the same data, delineating the space to be explored by more sophisticated end-to-end models	abstract	We introduce a new unsupervised task, spoken language modeling: the learning of linguistic representations from raw audio signals without any labels, along with the Zero Resource Speech Benchmark 2021: a suite of 4 black-box, zero-shot metrics probing for the quality of the learned models at 4 linguistic levels: phonetics, lexicon, syntax and semantics. We present the results and analyses of a composite baseline made of the concatenation of three unsupervised systems: self-supervised contrastive representation learning (CPC), clustering (k-means) and language modeling (LSTM or BERT). The language models learn on the basis of the pseudo-text derived from clustering the learned representations. This simple pipeline shows better than chance performance on all four metrics, demonstrating the feasibility of spoken language modeling from raw speech. It also yields worse performance compared to text-based 'topline' systems trained on the same data, delineating the space to be explored by more sophisticated end-to-end models.		
	versions		versions			