

ラジオ音声を対象とした音声強調における擬似学習データ利用の検討

小林 彰夫[†]大和大学 情報学部[†]安 啓一[‡]筑波技術大学 産業技術学部[‡]

1 はじめに

ラジオ放送は、災害時のメディアとして重要とされているが、雑音やフェージングといった妨害を受けやすく、受信音声の品質が容易に低下することから、難聴者にとってきわめて聴取が困難になる場合がある。低品質なラジオ音声の品質は、音声強調手法により改善できる一方、異なる受信環境に応じて頑健な学習ができるかどうかについての検討はあまり行われていない。そこで本稿では、異なる受信点で収集した学習データに基づくラジオ放送の音声強調、および学習データを擬似的に生成した場合の音声強調の有効性について検討を行ったので報告する。

2 ラジオ放送の音声強調

2.1 ラジオの通信路と品質低下の要因

わが国のラジオ放送では、振幅変調 (AM) もしくは周波数変調 (FM) 方式が採用されている。AM による中波放送は、FM に比べて変復調回路が単純であり、遠距離まで伝搬することから、特に災害時のメディアとしての強みを持つ。一方、FM に比べると雑音等による品質の低下の影響を受けやすい。品質低下の要因は様々であるが、本稿では加算性ガウス雑音およびフェージングを主要因として取り上げる (図 1)。

2.2 音声強調手法と擬似学習データの活用

本稿では、筆者らの既報告 [1] と同様、雑音抑圧オートエンコーダーの一つである CleanUNet[2] を用いて音

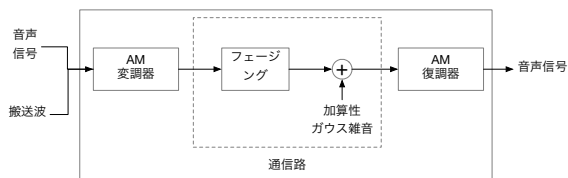


図 1: ラジオの通信路

声強調を行う。先の報告では、学習・評価に用いたデータの受信点が1か所であったことから、評価における受信環境はクローズドであり、受信環境がマッチしない場合の品質改善が不明であった。実用的な観点からは、未知の受信環境に対してもロバストな音声強調ができることが望ましい。そこで本稿では、複数の受信点で収集した低品質なラジオ音声の学習・評価データを用いて音声強調の頑健性を調べるとともに、伝搬上の品質低下要因に基づいて生成した擬似学習データによる音声強調について実験を行った。

3 実験

3.1 学習・評価データ

受信音声 本稿では、つくば市 (受信点 A) および大阪市 (受信点 B) の異なる2か所の受信点で受信したNHK ラジオ第1放送の音声を学習・評価データとした。両者とも、市販のループアンテナからSDR (Software Defined Radio) ドングルを経由して計算機上で録音を行った。ただし、アンテナと受信機との間に計算機を設置することにより故意に雑音を混入させ、音声品質を低下させた。また、中波放送の帯域は7.5kHzであることから、16kHzのサンプリング周波数を用いてラジオ音声を録音した。

リファレンス音声 NHK ネットラジオらじる★らじるから音声を収集してラジオの受信音声に対するリファレンスとした。リファレンスは16kHzにダウンサンプリングした。

擬似学習データ 2つのデータ拡大方法に基づいて、ネットラジオ音声の品質を低下させた学習用の音声データを生成した。ひとつは音声に対して加算性ガウス雑音の混合する拡大方法 (以下 GN)、もうひとつは GN と Clarke 法 [3] に基づくフェージングのシミュレーションを組み合わせた拡大方法 (以下 FD) である。

学習・評価データの構成 学習の際は、ラジオ音声とネットラジオ音声との相互相関を計算することによりデータ間のアラインメントを行い、ペアデータとした。

評価データは、受信点 A について471発話、受信点 B について500発話とした。品質の評価は客観評価指標で

A Study on Speech Enhancement for Degraded Radio Signals by Pseudo Training Data

[†]Akiyo KOBAYASHI

Faculty of Informatics, Yamato University

[‡]Keiichi YASU

Faculty of Industrial Technology, Tsukuba University of Technology

表 1: 受信環境の異なる学習データによる音声強調モデルの評価

評価データ		受信点 A			受信点 B		
学習データ		N/A	受信点 A	受信点 B	N/A	受信点 A	受信点 B
PESQ		1.084	1.568	1.148	1.071	1.114	1.222
DNSMOS	OVRL	1.414	2.094	1.533	1.463	1.775	1.710
	SIG	3.038	2.716	2.947	3.214	3.406	2.332
	BAK	1.272	2.822	1.546	1.277	1.618	2.400
HASQI	損失なし	0.181	0.320	0.146	0.070	0.109	0.156
	mild	0.245	0.391	0.161	0.100	0.144	0.169
	moderate	0.277	0.425	0.184	0.116	0.175	0.186

表 2: 擬似学習データによる音声強調モデルの評価

評価データ		受信点 A					受信点 B				
学習データ		N/A	受信点 A		受信点 B		N/A	受信点 A		受信点 B	
拡張法			GN	FD	GN	FD		GN	FD	GN	FD
PESQ		1.084	1.117	1.172	1.104	1.146	1.071	1.110	1.112	1.105	1.107
DNSMOS	OVRL	1.414	1.517	1.980	1.428	1.719	1.463	1.603	1.708	1.547	1.607
	SIG	3.038	2.707	3.623	2.839	3.535	3.214	3.319	3.535	3.288	3.498
	BAK	1.272	1.675	1.732	1.460	1.482	1.277	1.675	1.732	1.460	1.482
HASQI	損失なし	0.181	0.200	0.218	0.205	0.217	0.070	0.103	0.120	0.102	0.110
	mild	0.245	0.253	0.283	0.263	0.268	0.100	0.115	0.145	0.126	0.130
	moderate	0.277	0.287	0.331	0.294	0.313	0.116	0.135	0.172	0.148	0.154

ある PESQ, ITU-T P.835 の指標を non-intrusive に推定する DNSMOS P835[4], 聴力損失を反映した音声品質の客観指標である HASQI[5]を用いて行った. HASQI は健聴, 軽度難聴 (mild), 軽中度難聴 (moderate) の 3 つの聴力損失パターンで評価した.

3.2 実験結果

表 1 をみると, 学習・評価データの受信環境が一致する場合, PESQ, DNSMOS(OVRL), HASQI いずれも最良の値となった. 一方, DNSMOS(SIG) は必ずしも環境が一致した場合に最良の数値となっておらず, 背景雑音の抑圧などの効果が全体の評価 (OVRL) に影響しているのではないかと考えられる. 擬似学習データを用いたモデルの評価では (表 2), 実データで学習したモデルよりも各指標は低下する. また, 単に加法性ガウス雑音を追加するだけではなく, フェージングによる受信レベルの変動を加えた方が有効であることがわかった.

4 おわりに

ラジオを対象とした音声強調におけるモデルの頑健性および擬似学習データの利用について検討した. 実験結果から, 単純に加法性雑音を加えるよりも, フェージングのような受信レベルの変動を考慮した拡張方法が有効であることがわかった. その一方で実データを学習データとして用いた結果と比べると各種客観指標

の改善が小さかった. 今後は, 少量の受信音声から受信環境にマッチした擬似学習データを生成する手法について検討していきたい.

謝辞

本研究の一部は JSPS 科研費 20H01716, 21H00901, 23H00493, 23H00995 の助成を受けたものである.

参考文献

- [1] 小林ほか, “品質劣化したラジオ音声を対象とした音声強調手法の検討,” 情報処理学会全国大会, 2023.
- [2] Z. Kong et al., “Speech Denoising in the Waveform Domain with Self-Attention,” in *Proc. ICASSP*, 2022, pp. 7867–7871.
- [3] R. H. Clarke, “A Statistical Theory of Mobile-Radio Reception,” in *The Bell System Technical Journal*, 1968, pp. 957–1000.
- [4] C. K. Reddy et al., “DNSMOS P.835: A Non-Intrusive Perceptual Objective Speech Quality Metric to Evaluate Noise Suppressors”, in *Proc. ICASSP (ICASSP)*, 2022, pp. 886–890.
- [5] J. M. Kates et al., “An Overview of The HASPI and HASQI Metrics for Predicting Speech Intelligibility and Speech Quality for Normal Hearing, Hearing Loss, and Hearing Aids,” in *Hearing Research*, 2022.