

論文内容要旨 (和文)

2020 年度入学 大学院博士後期課程

電子情報工学 専攻

氏 名 中澤和司



論 文 題 目 参照信号を必要としない音声了解度推定方法に関する研究

音声了解度は音声の内容を正しく聞き取ることができることを表す指標であり、音響設計やシステムの評価に必要である。これらの指標を監視することで音声の聞き取り易さを保証することが可能になるが、被験者を用いた試聴試験から取得することは時間的コストがかかるため、これまでに音声了解度推定(SIP)アルゴリズムが開発されている。SIPには参照信号(クリーンな音声)を必要とするフルレファレンス(F-R)型と、参照信号を必要としないノンレファレンス(N-R)型がある。N-Rの推定方法はクリーンな音声が入らない状況でも予測をすることが可能であるが、F-R型より使用できる情報が少ないため推定が困難であるという欠点がある。この解決策として、音声強調(SE)の技術を利用したDeep Neural Network(DNN)ベースのN-R型SIPを提案した。

本研究では上記のSEを使用したモデルを参考に、より様々な歪に対応したN-R型SIPを開発すること目標として、残響劣化音声に対する了解度推定、さらに残響と加算雑音両方の劣化に対する推定が可能なモデルの作成を目指した。以下、本論文は以下の構成となっている。

1. 序論

音声了解度推定の必要性、これまでに提案されてきたF-R, N-R了解度推定方法の説明、それぞれの推定方法の欠点を述べる。本研究の意義について述べた。

2. 残響劣化音声に対する主観了解度測定

残響劣化音声に対応したSIP作成を目指して、被験者に対して残響劣化音声に対する主観了解度の測定を行った。主観了解度推定方法には日本語版DRT(JDRT)を使用した。4話者(男性2名、女性2名)の音声に対し室内インパルス応答(RIR)を畳み込むこと残響時間0.2から1.2秒の残響劣化音声生成し、7名の被験者に対して主観評価を行った。結果として了解度は残響時間の増加に伴い低下し、また一部の子音に対しては話者間で大きく了解度が異なることが示された。

3. 残響劣化音声に対するノンレファレンス了解度推定

2章で測定した残響劣化音声と主観了解度を使用して、モデルの学習と評価を行った。推定モデルに組み込むSE DNNの影響を確認するために、MetricGAN+とMask-MSELossという2つのモデルを使用した。了解度を推定するためのDNNの入力特徴量には(frequency-weighted segmental SNR)fwSNRsegを使用した。fwSNRseg演算中の帯域分割数 λ と重み付けパラメータ p を変えながら推定精度を行った結果、特徴量演算時のパラメータの影響は使用するSE DNNにより異なりMetricGAN+を使用した場合 $\lambda=64$ に $p=1$ としたときに最も高い精度で推定可能であった。また、既存のN-R型推定方法と比較しても高精度であることを示し、SE DNNを使用したSIPは残響劣化した音声にも有効に了解度推定可能であることを示した。

4. 残響劣化音声に対する学習可能な特徴量抽出フィルタを用いたノンレファレンス了解度推定

3章で使用した推定モデルでは、特徴量演算のパラメータ選択的に決めていた。本章では特徴量抽出部分をDNNに置き換えることでデータを元に最適化を行うことで推定精度向上を目指した。結果として提案手法で精度向上が可能であり、3章と同様にSE DNNにMetricGAN+使用時により高い精度で了解度推定可能であることを示した。また劣化音声だけを入力とした推定結果と比較することで、強調音声を使用することの有効性を示した。

5. 残響と加算により劣化した音声に対する主観了解度測定

多様な歪に対して了解度推定可能なモデルを作成するために、残響と加算ノイズにより劣化した音声に対する了解度推定を目指して主観評価値の測定を行った。主観評価方法には2章と同様にJDRTを使用して、残響の付加には2章で使用したRIRから0.3, 0.6, 0.9, 1.2秒の使用し、多人数の話し声から構成されるBabbleノイズを-10, -5, 0 dBで重畳した音声に対する主観評価を6名の被験者に対して行った。結果としては、SNRに関しては低下に伴い主観了解度が低下するが、残響時間に関しては大きく低下しないことが示された。

6. 残響と加算により劣化した音声に対するノンレファレンス了解度推定

2章、5章で測定したデータと、既存の加算雑音に対する主観評価データを用いて了解度推定モデルの学習と評価を行う。計算コスト削減のため公開されている学習済みSE DNNを使用して、4章で提案したモデルで学習と評価を行った。結果としてテストデータに対して既存の手法より高い精度で了解度推定が可能であることを示した。

7. 結論

本章では本研究の結論、新規性についてまとめた。また今後のSIPへの課題について述べた。

論文内容要旨 (英文)

2020

年度入学 大学院博士後期課程

電子情報工学

専攻

氏 名 _____ 中澤和司



論 文 題 目 A Speech Intelligibility Prediction Method Not Requiring the Reference Signal.

This study focuses on developing a Non-Reference (N-R) type Speech Intelligibility Prediction (SIP) model using Deep Neural Networks (DNNs) and Speech Enhancement (SE). The aim is to create a model that accurately estimates degraded speech intelligibility with both reverberation and additive noise distortions.

The paper begins by highlighting the significance of speech intelligibility prediction and discussing the limitations of conventional Full-Reference (F-R) methods which requires the reference signal, and N-R methods which do not require the reference. The importance of this research is emphasized.

To address these limitations, subjective intelligibility assessments are conducted for reverberant degraded speech. The Japanese Diagnostic Rhyme Test (JDRT) is employed for subjective evaluations, revealing a decrease in intelligibility with increased reverberation time and notable variations in intelligibility across different consonants.

Next, N-R type intelligibility prediction for reverberant degraded speech is explored using the collected data. Two models, MetricGAN+ and Mask-MSELoss are used for SE to estimate the reference signal. The frequency-weighted segmental signal-to-noise ratio (fwSNRseg) between this signal and the degraded signal is input to the SIP DNN. The results indicate that the choice of parameters in fwSNRseg computation depends on the SE DNN used, and the proposed SE DNN-based SIP outperforms existing N-R methods, even for degraded reverberant speech.

Furthermore, a DNN-based feature extraction filter is introduced to optimize feature computation, resulting in improved prediction accuracy. The effectiveness of using enhanced speech is demonstrated by comparing the results with those obtained solely from degraded speech inputs.

The study is further extended to evaluate intelligibility for both reverberation and additive noise distortions. Subjective evaluations using JDRT are conducted for speech samples degraded with various levels of reverberation and babble noise. The results show that intelligibility decreases with decreasing SNR, but the impact of reverberation time on intelligibility is not as significant.

In conclusion, an N-R SIP model that can accurately predict intelligibility with various distortions, including reverberation and additive noise was proposed. The findings highlight the effectiveness of SE DNN-based approaches for estimating intelligibility and suggest future research directions for SIP.

学位論文の審査及び最終試験の結果の要旨

令和 5 年 8 月 1 日

理 工 学 研 究 科 長 殿

課程博士論文審査委員会

主査 近藤和弘

副査 小坂哲夫

副査 湯浅哲也

副査 安田宗樹

副査 大槻恭士

学位論文の審査及び最終試験の結果を下記のとおり報告します。

記

論 文 申 請 者	専攻名 電子情報工学専攻			氏名 中澤和司
論 文 題 目	参照信号を必要としない音声了解度推定方法に関する研究			
学位論文審査結果	合格	論文審査年月日	令和 5 年 7 月 18 日～ 令和 5 年 8 月 1 日	
論文公聴会	令和 5 年 8 月 1 日	場 所	工学部 7 号館 301 教室	
最終試験結果	合格	最終試験年月日	令和 5 年 8 月 1 日	

学位論文の審査結果の要旨 (1,000 字程度)

本論文では残響による劣化のある音声の了解度を、原音を用いずに推定する方法について検討し、評価している。現在では様々な環境下で音声通信が行われており、音声品質を実時間でモニタリングすることが望ましい。このモニタリングは受信端で行うことが想定されるが、通常受信端では劣化音声しか入手できない。しかし、一般的に品質の評価は原音と劣化音声を比較することで行われる。よって実現可能なモニタリングのためには、劣化音声のみから通信品質を推定することが必要であるが、本論文では深層学習(DNN)を用いて原音を推定する方法を提案してこれを実現することを試みている。更にこの推定原音と劣化音声の差分から音声了解度を DNN により推定する方法を検討した。その結果、十分実用的な推定精度が実現できることを示した。対象とする劣化も、加算雑音に加え、残響に対しても十分な推定性能を発揮できることを示した。その推定精度は類似方式と同等、あるいはそれ以上であることも検証している。以上、本研究のように詳細評価を系統的に行い、その実用性を証明した研究は類を見ない。

本学位論文は研究背景、目的とも十分練られており、現状の問題点を指摘しこれを解決する方法について独自の工夫を述べている。本編も詳細にその提案推定手法の説明、およびその性能評価手法について説明されており、その実験結果、ならびに考察も十分に述べられている。最終的には提案手法により多様な劣化を含む音声信号に対し実用に耐えうる精度で了解度推定が可能であることを示している。最後に本研究の新規性、貢献を詳細に述べ、今後さらに解決すべき課題も挙げている。

本研究成果は、了解度推定法に独自の工夫があり、新規性、工学的応用として十分貢献できると認められる。この成果は 1 編の学術論文、8 編の国際会議論文（うち海外渡航で 3 編）としてその内容を公開しているうえに、多数の国内学会、研究会で発表を行っている。さらに発表論文 2 件が優秀発表賞を受賞するにも至っている。以上により、本学位論文は工学的貢献が十分にあると認められるため、博士学位論文として合格と判定した。

本論文の内、主観評価に関する項目については工学部倫理審査委員会の許可（承認番号 R4 - 10）を得て実施している。また、本論文は利益相反等に係る学内規則に基づく手続きは必要ない。

最終試験の結果の要旨

博士論文公聴会における発表の様子、質疑応答の対応を通じて、研究目的の設定、研究の進め方、問題発掘および解決能力、関連知識の豊富さ、理解力などを検査し、博士（工学）に十分値する知識と能力があると認められる。さらに国際学会発表における英語発表、ならびに質疑応答の対応の様子、および公聴会後の口頭試問から、博士（工学）に求められる語学力やプレゼンテーション・スキルも十分あるものと認められたので、合格と判定する。