

学位論文題名

A study of high quality speech synthesis based
on the analysis of the randomness in speech signals

(音声信号におけるランダムネスの解析に基づいた
高品質音声合成に関する研究)

学位論文内容の要旨

近年、マン・マシンインタフェースのマルチメディア化が精力的に進められており、その一翼を担う音声合成技術の分野では、より肉声らしく聞こえる高品質な合成音声を生成する手法の開発が目下の課題とされている。本論文は、ヒトの音声信号に観察されるいくつかの種類のランダムネスについて解析を行い、その結果に基づいて人工的に生成したランダムネスを合成音声に付加することで合成音声の自然性が向上することを明らかにした、音声合成および音声知覚に関する基礎的な研究の結果をまとめたものである。研究の対象としたランダムネスは、母音音声信号における(1)振幅ゆらぎ、(2)ピッチゆらぎ、(3)波形ゆらぎ、および(4)線形予測分析より求めた音源信号に固有のランダムネス、(5)有声子音に見られる高周波成分の非周期性の計5種類とした。

本論文は、まず第二章で振幅ゆらぎおよびピッチゆらぎの統計的性質および音声知覚に及ぼす影響について述べている。解析の結果、振幅ゆらぎ、ピッチゆらぎはその周波数特性がおおよそ $1/f$ 特性となる雑音性のゆらぎ系列であることがわかった。また、それぞれのゆらぎを白色雑音としてモデル化するとき合成音声は粗雑な病的音声を思わせる品質となり、その一方で $1/f$ 雑音とすると健康者の音声音質に近づき自然性が改善されることがわかった。結果として、(1)振幅ゆらぎ、ピッチゆらぎの周波数特性は合成音声の自然性に影響を及ぼす主要因の一つであること、(2)両者をランダムフラクタル($1/f$ 雑音)としてモデル化するとき自然性が改善されることがわかった。

第三章では波形ゆらぎの統計的性質および音声知覚に及ぼす影響について述べている。解析の結果、母音音声信号から線形予測分析により得られた残差信号を二階積分した周期的信号におけるピッチ周期ごとの波形変化は、その周波数特性がおおよそ $1/f^2$ 特性となる雑音性の信号であることがわかった。また、こうした波形ゆらぎを白色雑音としてモデル化するとき合成音声は音声と背景雑音に分離して知覚され、その一方で $1/f^2$ 雑音とすると健康者の音声音質に近づき自然性が改善されることがわかった。結果として、(1)波形ゆらぎの周波数特性もまた振幅ゆらぎおよびピッチゆらぎと同様、音声の自然性に影響を及ぼす主要因の一つであること、(2)波形ゆらぎをランダムフラクタル($1/f^2$ 雑音)としてモデル化するとき自然性が改善されることがわかった。

第四章では音源信号に固有のランダムネスについて述べている。線形予測分析により得られた残差信号は白色特性であることが仮定されるため、これを一階積分することで得られる線形予測合成

フィルタの音源信号には $1/f^2$ 雑音の性質が現れることが予想できる。本論文では、シャウダー展開を用いて多重解像度解析を行い、この仮説の検証を試みた。その結果、(1) 健常者が発声した実際の音声から逆フィルタリング処理により求めた音源信号は、人工的な音源信号として頻繁に使用されるインパルス列音源とは異なり、波形の不連続性が顕著には見られないこと、(2) 音源信号を観察する解像度が大きくなるにつれ、 $1/f^2$ 雑音としての性質が顕著になることがわかった。これらの知見に基づき、健常者の音源信号と同じ統計的性質をインパルス列音源に反映させる方法を新たに提案し、音響心理実験を通して提案法の有効性を確認した。提案法は、インパルス列音源により生成された合成音声で特に問題とされるブザー音質化の改善に効果があることがわかった。

第五章では有声子音に見られる高周波成分の非周期性の特徴および音声知覚に及ぼす影響について述べている。有声子音では非周期的な無声区間（子音区間）とそれに続く周期的な有声区間（母音区間）の境界が判然としないため、有声区間であっても特に高周波成分では非周期成分が顕著に表れる傾向を示す。このような音声の性質を利用した符号化方式に MELP (Mixed Excitation Linear Prediction) 方式があり、近年、新しい低ビットレート符号化方式の標準の一つとして利用が進んでいる。本研究ではウェーブレット変換を用いた MELP 方式の構成方法を提案し、サブバンド分解後の各帯域での有声・無声（周期性・非周期性）判定から有声子音の統計的性質について解析を試みた。また、解析結果に基づき、主として高周波数域のサブバンドを非周期性成分の支配的な帯域として合成した有声子音では、了解度および自然性が向上することがわかった。第五章ではさらに、本研究でこれまでに明らかにした特徴をすべて取り入れた実時間動作可能な音声符号化合成システムおよび規則合成システムをそれぞれ DSP (Digital Signal Processor) 評価ボードおよび汎用パーソナルコンピュータ上で実現し、提案手法の有効性を具体的に実証した結果についても言及している。構築したそれぞれの音声合成システムは、有声音および無声音の合成にそれぞれインパルス列と白色雑音を音源信号として使用する従来の音声合成システムよりも、より肉声らしく健常者の音声音質を思わせる合成音声を生成することが可能であることがわかり、結果として、一連の提案手法は合成音声の品質の改善に有効であることがわかった。

結論として、本論文では、(1) ヒトの音声信号に観察されるランダムネスを音声合成の際に適切に再現することで、合成音声の品質が向上すること、(2) 本研究により得られた知見に基づいて開発された提案手法は、従来の音声合成システムに簡単に組み込むことが可能であり、高品質音声合成システムを実現するための具体的なノウハウとして実際に利用が可能であることを明らかにした。本論文ではさらに、(3) 研究対象としたいいくつかのランダムネスの場合にはランダムフラクタルとしての性質が観察され、このような音声信号のランダムフラクタル性がヒトの音声信号の特徴を定量的に調べる上で、一つの有用な観点となり得る可能性があることを明らかにするとともに、こうしたランダムフラクタル性の再現が合成音声の自然性改善に優れた効果があることを、音声合成技術の分野における新たな知見として明らかにした。

学位論文審査の要旨

主 査 教 授 伊 福 部 達
副 査 教 授 河 原 剛 一
副 査 教 授 清 水 孝 一

学 位 論 文 題 名

A study of high quality speech synthesis based on the analysis of the randomness in speech signals

(音声信号におけるランダムネスの解析に基づいた
高品質音声合成に関する研究)

近年、マン・マシンインタフェースのマルチメディア化が精力的に進められており、その一翼を担う音声合成技術の分野では、より肉声らしく聞こえる高品質な合成音声を生成する手法の開発が目下の課題とされている。本論文は、ヒトの音声信号に観察されるいくつかの種類のランダムネスについて解析を行い、その結果に基づいて人工的に生成したランダムネスを合成音声に付加することで合成音声の自然性が向上することを明らかにした、音声合成および音声知覚に関する基礎的な研究の結果をまとめたものである。研究の対象としたランダムネスは、持続発声母音における(1) 振幅ゆらぎ、(2) ピッチゆらぎ、(3) 波形ゆらぎ、および(4) 線形予測分析によって求めた音源信号のランダムフラクタル性、(5) 有声子音に見られる高周波成分の非周期性の計5種類とした。

本論文は、まず第二章で振幅ゆらぎおよびピッチゆらぎの統計的性質および音声知覚に及ぼす影響について述べている。結果として、これらのゆらぎは合成音声の自然性に影響を及ぼす要因の一つであること、これらをランダムフラクタル ($1/f$ 雑音) としてモデル化するとき自然性が改善される可能性があることを明らかにした。

第三章では波形ゆらぎの統計的性質および音声知覚に及ぼす影響について述べている。結果として、波形ゆらぎもまた音声の自然性に影響を及ぼす要因の一つであり、これをランダムフラクタル ($1/f^2$ 雑音) としてモデル化するとき自然性が改善される可能性があることを明らかにした。

第四章では線形予測分析によって得られた音源信号に見られるランダムフラクタル性について述べている。結果として、健常者の持続発声母音から求めた音源信号は、人工的な音源信号として頻繁に使用されるインパルス列音源とは異なり、波形の不連続性が顕著ではないこと、音源信号を観察する解像度が大きくなるにつれて $1/f^2$ 雑音としての性質が見られるようになることがわかった。これらの知見を利用して、本章ではインパルス列音源により生成された合成音声で特に問題とされるブザー音質化を改善する方法を新たに提案し、その有効性を音響心理実験によって明らかにした。

第五章では、前章までに得られた知見に基づいて設計された実時間動作可能な音声符号化合成シ

システムおよび規則合成システムの開発について述べている。また、有声子音の自然性を改善することを目的として、ウェーブレット変換を用いた MELP (Mixed Excitation Linear Prediction)方式の構成方法を新たに提案した。構築したそれぞれの音声合成システムは、有声音および無声音の合成にそれぞれインパルス列と白色雑音を音源信号として使用する従来の音声合成システムよりも、より肉声らしく健常者の音声音質を思わせる合成音声を生成することが可能であり、結果として、本論文で提案した自然性改善のための一連の方法は合成音声の自然性の改善に効果があることがわかった。

結論として、本論文では、(1) ヒトの音声信号に観察されるランダムネスを音声合成の際に適切に再現することで、合成音声の品質が向上すること、(2) 本論文で提案した自然性改善の方法は、高品質音声合成システムを実現するための具体的なノウハウとして実際に利用が可能であることを明らかにした。本論文ではさらに、(3) 研究対象としたいいくつかのランダムネスの場合にはランダムフラクタルとしての性質が観察され、このような音声信号のランダムフラクタル性がヒトの音声信号の特徴を定量的に調べる上で、一つの有用な観点となり得る可能性があることを明らかにするとともに、こうしたランダムフラクタル性の再現が合成音声の自然性改善に優れた効果があることを、音声合成技術の分野における新たな知見として明らかにした。

これを要するに、著者は、ヒトの音声信号に観察されるいくつかのランダムネスについてその諸性質を明らかにし、合成音声の自然性改善に関する新知見を得たものであり、音声信号処理工学の分野に対して貢献するところ大なるものがある。

よって著者は、北海道大学博士（工学）の学位を授与される資格あるものと認める。