

平成 21 年 4 月 16 日現在

研究種目：若手研究(B)

研究期間：2006～2008

課題番号：18700171

研究課題名（和文）非定常音源モデルに基づく高品位音声合成システムの開発

研究課題名（英文）Development of high quality speech synthesis system based on unsteady glottal source model

研究代表者

野村 英之 (NOMURA HIDEYUKI)

電気通信大学・電気通信学部・助教

研究者番号：90334763

研究成果の概要：自然性を有する音声合成システム開発のため、音源に含まれるゆらぎをシミュレーションにより検討する。音声には振幅、基本周波数、及び波形のゆらぎが含まれ、肺圧の増加により増加した。さらに、左右非対称な声帯モデルによるシミュレーションとゆらぎ評価を行ない、収録音声のゆらぎと比較を行った。対称モデルによるゆらぎと、収録音声に含まれるゆらぎは一致した。非対称モデルによる音声のゆらぎは収録音声のものより増加した。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2006 年度	1,300,000	0	1,300,000
2007 年度	1,100,000	0	1,100,000
2008 年度	700,000	210,000	910,000
年度			
年度			
総計	3,100,000	210,000	3,310,000

研究分野：総合領域

科研費の分科・細目：知覚情報処理・知能ロボティクス

キーワード：音声情報処理，声帯音源モデル，非定常音源，ゆらぎ

1. 研究開始当初の背景

社会の情報化にともない、身の周りの多くの機器のインテリジェント化が進み、従来では考えられなかった多くの機能が組み込まれている。

また、これまで工場での作業支援が多くの活躍の場であったロボットが社会生活支援を目的とし、通常の生活でも使用が望まれている。これらのロボットの最終目標は「鉄腕アトム」であろうが、現時点ではそのレベルに達していない。

こういったロボットを含めたインテリジェント機器が日常生活により深く定着するには、それが機械であることを意識しない操作方法、すなわちインターフェースが必要となってくる。そのため

には人間が普段から使用するコミュニケーション手段を利用することが望ましい。その代表が音声を利用する手段である。このためには音声認識や音声合成といった技術開発が必要となってくる。

現在、音声認識は雑音のない理想的な条件ではかなり実用化が進められている。音声合成も単に人間の言葉を発声(合成)するだけなら、ほぼ完成している技術である。一方で人間の感情を読み取る、また、感情をこめた声の合成には未解決な問題が多い。単に自然性のある、人間の音声を合成することですら、難しいのが現状である。このような背景から、人間らしい発声

する技術の開発が望まれている。

2. 研究の目的

本研究は「インテリジェント機器の音声インターフェース及びロボットのための音声対話システム」を開発する研究の一環で実施される。特により自然性を有する発声機構の構築をめざし、音声合成技術において、従来考慮されていなかった「音源の非定常性」に着目し、機械的で低品質な合成音に「より自然性」をもたせるための基礎的検討を行う。

合成音声那不自然である理由は、音声生成過程を忠実に模擬していないことがひとつの理由である。合成音声の音源信号は、周期的なパルス列を用いることが多い。ところが、実際の音声生成過程の音源信号(断続的な声門流)は非定常な振る舞いを示す。

本研究では、この音源の非定常な成分として「振幅ゆらぎ」、「ピッチゆらぎ」、及び「波形ゆらぎ」に着目し、音源特性の解明を目的とする。特に(1)非定常性を考慮した数値音源モデルの構築、(2)提案数値音源モデルによるシミュレーション、及び(3)シミュレーション音声のゆらぎ評価、の3項目の研究を実施することを目的とする。さらに提案モデルの応用として、(4)疾患音声の合成シミュレーションを試みる。

3. 研究の方法

本研究では合成音声用音源に、実際の人間の音声に含まれるゆらぎ成分を付加するための基礎検討として、シミュレーションによる検討を中心に行う。また、疾患音声のシミュレーション、及び実収録音声との比較を行う。

(1) 声帯音源数値モデルの構築

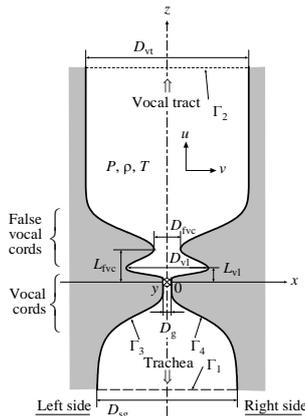


図1 喉頭モデル。

図1に提案するモデルを示す。モデルは音源に相当する声帯(Vocal cords)、共鳴管に相当する声道(Vocal tract)、及びリザーバタンクに相当する気管(Trachea)から構成される。このモデルにおける音声生成は以下のとおりである。

① まず、モデル下端に肺圧に相当する圧力

を加える。

- ② 声道終端部分との圧力差により、声門(声帯で形成される狭め)を通過する空気流、すなわち声門流れが発生する。
- ③ この声門流れが声門の狭めを上げようとする力を与え、声門が開く。
- ④ 声門が拡大することで、肺圧と声道終端との圧力差が減少し、声門流の速度が減少する。
- ⑤ 速度減少により、声門の間隔が狭まる。

この①から⑤の繰り返しにより、声帯音源といわれる、準周期的な流れが生成される。この音源が声道の共鳴周波数により、音色が付加され、音声(母音)が生成される。

この研究では、この減少を流体力学の支配方程式に基づきシミュレーションを行う。シミュレーションでは時々刻々と変化する圧力や流速を計算する。

③から⑤の課程で声門流により、声帯が振動する。このことを模擬するため、声帯を等価的な機械要素で表現する。そのモデルが図2である。声帯の有効質量を有する弾性膜(Elastic cover)と、その機械的特徴を決定する分布した無数のばね(Spring)とダンパ(Damper)で構成する。これらの機械要素は声門流の圧力と結合し、その共振周波数で振動する。

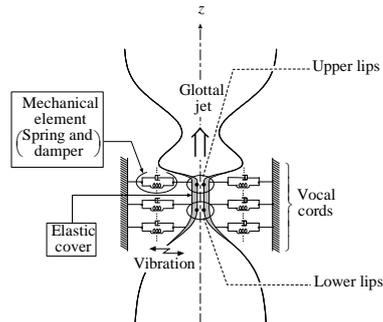


図2 声帯の機械モデル。

(2) 提案モデルによる音声生成シミュレーション

(1)で提案した物理モデル(喉頭モデル、及び声帯モデル)を偏微分方程式で表現し、その振る舞いを数値的にシミュレーションする。声門流れは流体のもつ粘性及び熱伝導性を考慮した式であり、これを有限差分法で解く。声帯の振動は連立させた運動方程式で表現し、Runge-Kutta法で解く。両者を計算ステップ毎に解き、音声生成を模擬する。

シミュレーションに用いるプログラムはC++言語で開発する。また、数値計算はワークステーションを利用する。

シミュレーションは肺圧(Lung pressure) P_{L0} を加圧する。声道終端は音波の反射のない無反射終端条件を課す。喉頭、及び声帯の形状と力学的な特性は他文献による実測値、または推定値を利用する。ただしシミュレーションで想定す

る声道形状は幅一様とし、日本語母音の/e/に近いものとする。

(3) ゆらぎ評価

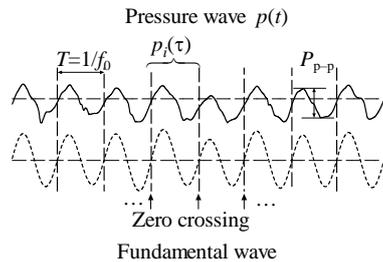


図3 母音波形の分割化。

母音波形をピッチごとに切り出し(図3参照), 生成音の揺らぎ評価として, 振幅, ピッチ(基本周波数), 及び波形の3種類の揺らぎを評価する。

- ① 振幅揺らぎ: ピッチごとに抽出した波形の個々の振幅の変動を評価する。評価は抽出した振幅の平均値 m と標準偏差 σ の比である変動係数 $CV = \sigma/m$ で行う。
- ② 基本周波数揺らぎ: ピッチごとの波形の基本周波数の変動係数, すなわちその平均値と標準偏差の比を評価する。
- ③ 波形揺らぎ: ピッチごとの波形の統計平均を平均波とし, そのエネルギーを H とする。平均波と各ピッチの波形との差を揺らぎ成分とし, 平均エネルギーを N とする。波形揺らぎは Harmonic-to-noise ratio, $HNR = H/N$ で評価する。

(4) 疾患音声のシミュレーションと揺らぎ評価

ポリープや腫瘍を有する喉頭疾患の場合, 生成音は健常者と比較し, ゆらぎが大きくなると予想される。そのことを確認, 検討するため, 疾患音声の生成シミュレーションを行う。ここでは左右声帯の力学的特性, 及び幾何学的特性が異なり, 非対称な声帯における音声生成シミュレーション, また揺らぎ評価を行う。

さらに健常者である男性3名の日本語5母音を収録し, その揺らぎと比較することで, 疾患の程度と生成音の揺らぎの評価を行う。

4. 研究成果

(1) シミュレーション例

声帯の特性が左右対称かつ, 健常者のそれに合わせてシミュレーションを行った。図4にその一例を示す。(a), (b), (c)はそれぞれ声門から下流方向(口唇方向, 図1の上方向)へ距離0, 20, 160 mmの位置(4段目の図中の○に対応)での圧力波形 p である。なお音圧は肺圧 $P_{L0} = 800$ Pa (ほぼ通常発生時の値)で規格化されている。

この結果から, いずれの観測位置においても圧力波形は振幅, および周期(基本周波数)が

一定ではなく, 変動(ゆらぎ)が生じていることが確認できる。また, ピッチごとの波形自身もゆらいでいることがわかる。このように, 声帯に疾患がないにも関わらず, 生成される音声には揺らぎが含まれる。これらの揺らぎは, 声門流れ自身が層流ではなく, 時間及び空間的に変動する乱流的な振舞いをするためである。

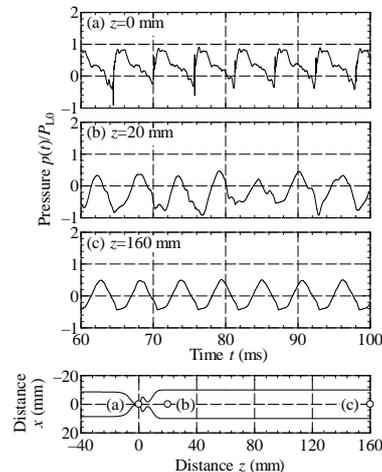


図4 シミュレーション波形。

(2) 肺圧による揺らぎ変化

音声生成シミュレーション肺圧 P_{L0} を変化させながら行い, ゆらぎの評価を行った。その結果が図5である。図の横軸は肺圧, 縦軸は上から基本周波数 f_0 の変動率(CV of f_0), 振幅 P_{p-p} の変動率(CV of P_{p-p}), 及び波形揺らぎ(HNR)である。

この結果は, 肺圧の上昇により, いずれの揺らぎも増加することを示す。また音源よりわずかに下流(20 mm)での揺らぎが最も大きいことがわかる。

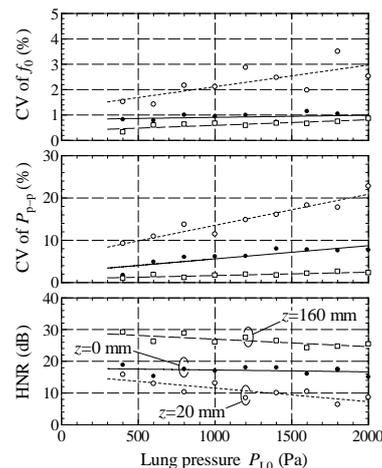


図5 肺圧と揺らぎの関係。

(3) 疾患音声シミュレーション揺らぎ評価

健常者の場合, 左右声帯の(幾何学的, 及び力学的)特性はほぼ対称であると考えられる。そこで, その特性を左右非対称にした喉頭疾患モ

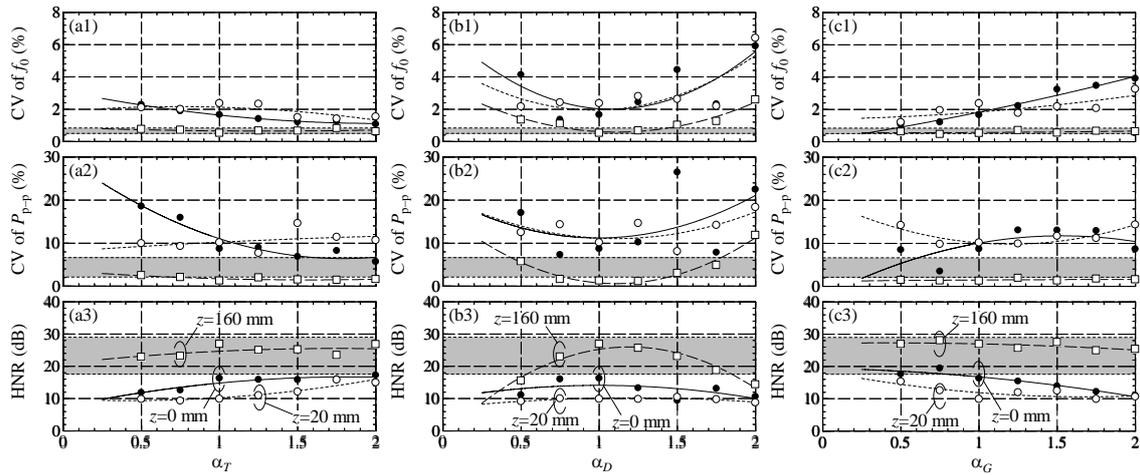


図6 幾何学的非対称性とゆらぎの関係。

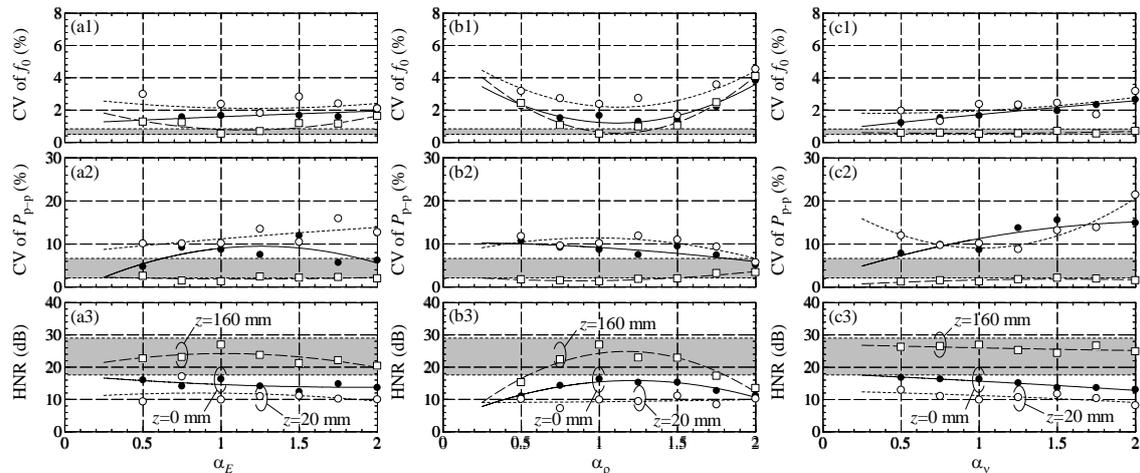


図7 力学的非対称性とゆらぎの関係。

デルにおいて、疾患音声のシミュレーション、及びゆらぎ評価を行った。

図6は幾何学的特性の非対称性と、得られたシミュレーション音声のゆらぎの関係である。左から声帯の厚さ(a1, a2, a3), 声帯の有効部分の深さ(b1, b2, b3), 及び喉頭中心からの声帯表面までの距離(c1, c2, c3)の非対称性とゆらぎの関係である。また、上から基本周波数の変動係数(CV of f_0), 振幅の変動係数(CV of P_{p-p}), 及び波形ゆらぎ(HNR)である。なお、横軸は左側声帯の特性を標準値に固定し、右側声帯の特性を変化させた場合の値の比である。また、グレーの部分は実収録音声解析して得られたゆらぎである(平均値±標準偏差)。

一般に、左右声帯の非対称性が増加すると、3種類のゆらぎが増加する結果となった。特に、声帯の有効振動部分の深さの非対称性(b1, b2, b3)の変化が大きく影響している。これは、他の特性は単純に形状自身に影響するのに対し、声帯の深さ変化は、声帯の有効質量、弾性率、及び粘性率のいずれにも影響、そしてその共振周波数を変化させるためである。

実収録音声のゆらぎは左右対称モデル(横軸=0)の160 mm位置での値とほぼ一致する。非対

称性が増加することで、HNRの値が実収録音声のゆらぎ範囲から外れる結果である。

図7は力学的特性の非対称性と、得られたシミュレーション音声のゆらぎの関係である。左から弾性率(a1, a2, a3), 密度(b1, b2, b3), 及び粘性率(c1, c2, c3)の非対称性とゆらぎの関係である。データの表記方法は図6と同じである。

幾何学的特性の場合と同様に、非対称性の増加で各ゆらぎは増加する。3種類の非対称性の中では密度の非対称性が最もゆらぎに影響を与える。

実収録音声のゆらぎとは、左右対称モデルの160 mmの位置でのゆらぎとほぼ一致した。HNRに関しては非対称性の増加により、実収録音声のゆらぎ範囲から外れる結果となった。

(4) まとめ

より自然性を有する合成音声システム開発として、ゆらぎを考慮した音源の検討を行った。その基礎検討として、声帯音源の数値モデルを構築し、計算機シミュレーション、及びシミュレーション音声のゆらぎ評価を行った。

左右声帯の特性が対称な、健常者を想定した声帯音源モデルによるシミュレーション音声にさえ、基本周波数、振幅及び波形にゆらぎが含

まれていた。特に、肺圧を上昇させることで、そのゆらぎは増加した。このことから、合成音声の音源にゆらぎを与えることは重要な点だといえる。

提案モデルの応用として、疾患音声の生成シミュレーション、及びゆらぎ評価を行い、さらに収録音声のゆらぎと比較を行った。声帯のパラメータの非対称性が増加することで、シミュレーション音声のゆらぎは増加した。また、収録音声のゆらぎと比較を行ったところ、非対称性の増加により、シミュレーション音声のひずみは、収録音声のそれより大きくなった。このことは、提案した音源モデルが合成音声システムの発展はもとより、疾患音声の解析等の医用工学分野でも大いに応用できることを示唆する。

今後の課題として、音源のゆらぎをどのように合成音声システムへ組み込むかの検討が必要となってくる。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 3 件)

- ① H. Nomura and T. Funada, “Fluctuations in pressure wave within the larynx,” *Acoust. Sci. & Tech.*, **30**, 55–58 (2009) (査読有).
- ② H. Nomura and T. Funada, “Effects of the false vocal folds on sound generation by an unsteady glottal jet through rigid wall model of the larynx,” *Acoust. Sci. & Tech.*, **28**, 403–412 (2007) (査読有).
- ③ H. Nomura and T. Funada, “Sound generation by unsteady flow ejecting from the vibrating glottis based on a distributed parameter model of the vocal cords,” *Acoust. Sci. & Tech.*, **28**, 392–402 (2007) (査読有).

[学会発表] (計 13 件)

- ① H. Nomura, T. Kamakura, and T. Funada, “A numerical analysis of fluctuations in pressure wave within the larynx using two-dimensional asymmetrical vocal folds model,” in *Proc. Acoustics '08*, pp. 2375–2380 (2008. 7. 1, Paris, France).
- ② 野村英之, 船田哲男, “非対称声帯モデルによる音声生成シミュレーション,” 日本音響学会 2007 年秋季研究発表会講演論文集 2-4-18, 481–482 (2007. 9. 20, 山梨大).
- ③ H. Nomura and T. Funada, “A numerical experiment on waveform fluctuation within the larynx,” in *Proc. 19th International Congress on Acoustics, ICA 2007* (on CD-ROM, 2007. 9. 4, Madrid, Spain).
- ④ H. Nomura and T. Funada, “Effect of unsteady glottal flow on the speech

production process,” in *Proc. Interspeech 2007, 10th European Conference on Speech Communication and Technology*, pp. 1390–1393 (2007. 8. 29, Antwerp, Belgium).

- ⑤ 野村英之, 船田哲男, “非対称声帯モデルによる疾患音声生成の数値シミュレーション,” 電子情報通信学会技術研究報告 SP2007-43 (2007. 7. 27, 富山県立大).
- ⑥ H. Nomura and T. Funada, “Amplitude fluctuation of pressure wave within the larynx,” in *Proc. the Japan-China Joint Conference on Acoustics, JCA 2007* (on CD-ROM, 2007. 6. 4, Sendai, Japan).
- ⑦ 野村英之, 船田哲男, “声門ジェット流れの非定常化の検討,” 日本音響学会 2007 年春季研究発表講演論文集 1-Q-26, 315–316 (2007. 3. 13, 芝浦工大).
- ⑧ H. Nomura and T. Funada, “Lung pressure dependence of glottal sound source (A),” 4th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan (2006. 12. 2, Honolulu, Hawaii), *J. Acoust. Soc. Am.*, **120**, 3372 (2006).
- ⑨ 西村仁, 野村英之, 船田哲男, “非対称声帯モデルによる喉頭疾患の声帯振動解析,” 平成 18 年度電気関係学会北陸支部連合大会講演論文集 B-12 (2006. 9. 16, 金沢工大).
- ⑩ 西村仁, 野村英之, 船田哲男, “喉頭疾患時における声帯振動の数値解析,” 日本音響学会 2006 年秋季研究発表会講演論文集 1-6-2, 157–158 (2006. 9. 13, 金沢大).
- ⑪ 野村英之, 船田哲男, “声門で発生する乱流音源に与える仮声帯の影響,” 電子情報通信学会技術研究報告 SP2006-30 (2006. 7. 21).
- ⑫ H. Nomura and T. Funada, “Numerical experiment on lung pressure dependence of glottal sound source,” in *Proc. 5th International Conference of Voice Physiology and Biomechanics, ICVPB 2006*, pp. 57–60 (2006. 7. 12, Tokyo, Japan).
- ⑬ H. Nomura and T. Funada, “Dependence of vocal cords vibration on lung pressure: Numerical modeling and experiment,” in *Proc. 9th Western Pacific Acoustic Conference, WESPAC IX 2006* (on CD-ROM, 2006. 6. 26, Seoul, Korea).

[その他]

ホームページ等

6. 研究組織

(1) 研究代表者

野村 英之 (NOMURA HIDEYUKI)

電気通信大学・電気通信学部・助教
研究者番号:90334763

(2) 研究分担者
()

研究者番号:

(3) 連携研究者
()

研究者番号: