成蹊大学理工学研究報告
J. Fac. Sci.Tech., Seikei Univ.
Vol.52 No.2 (2015)pp.5-10

株式市況音声合成システム

世木　寛之[*1]

An Automatic Broadcast System Using Speech Synthesis for the Stock Market Report

Hiroyuki SEGI[*1]

ABSTRACT : The 'Kabushiki Shikyo' program broadcast on NHK Radio 2 reports on the daily closing prices and net changes of about 830 stocks listed on the Tokyo Stock Exchange. Reading out the numerical values without making mistakes within the allotted broadcast time can be extremely difficult for the announcers. We have therefore developed an automatic broadcast system for stock-price bulletins, which uses numerical speech synthesis and automatic speech-rate conversion. Our system has been used in experimental digital terrestrial radio broadcasts since October 2006 and also used in NHK radio 2 since March 2010. This article describes the generation of texts to build the speech waveform database, the mechanism used to synthesize numerical speech via the database, and the evaluation of naturalness for the synthesized speech samples.

Keywords : text-to-speech, speech rate conversion, stock market report, broadcasting, concatenative speech synthesis

(Received October 5, 2015)

1．はじめに

NHKの「株式市況」は，1925 年 3 月 23 日からラジオで放送を開始し[1]，現在ではラジオ第 2 放送で平日 17 時から放送されている超長寿番組である。放送時間は約 1 時間で，主に東証一部に上場している銘柄のうちの約 800 銘柄の終値と前日比を伝えている。1 時間近くもの間，連続して 1 人のアナウンサーが読み続けることは難しいため，2 人のアナウンサーが途中で交代して読み上げていた。それでも，1)数値を間違えずに読むこと，2)限られた時間内に収まるように，読みながら時間配分を調整することが要求され，アナウンサーにとって非常に難度の高い業務であった。

このため，2003 年秋頃からNHK放送技術研究所(以下技研)で，株式市況の音声合成化の検討を行い，細かな経緯は 3 章で述べるが，2010 年 3 月 29 日から音声合成による放送を開始した。また，2014 年 3 月からは 2 代目の株式市況音声合成システムがリリースされることになった。本報告では，研究・開発を行った株式市況の音声合成システムのしくみについて紹介する。

[*1] : 情報科学科准教授 (segi@st.seikei.ac.jp)

2．放送用音声合成の歴史と本システムの特徴

音声合成は 20 年ほど前から放送で利用されるようになった[2]。読み手（アナウンサー）がいなくても情報が伝えられることは，放送局にとって非常に利便性が高いからである。

放送局で使われてきた音声合成システムとしては，天気予報の自動送出[2-5]，交通情報[6]，スポンサー名テロップの読み上げ[7]が挙げられる。また，文献としては残っていないが，実際に使用された音声合成システムも数多くあり，例えば，テレビ埼玉では 1998 年 4 月 1 日から現在まで天気予報を音声合成で放送している。また，「株式市況」における株価情報の一部（銘柄名，株価の引け値，前日比）は，1990 年台半ばから 2004 年頃まで(正確な日時は不明)，ラジオNIKKEI(当時のラジオたんぱ)にて放送された。さらに，「いつでもニュース」は，2003 年 10 月 10 日から 2007 年 3 月 20 日まで，NHKデジタルラジ

オ実用化試験放送で放送された。この番組では，NHKのホームページに掲載されたニュース文を，波形編集の音声合成ソフトで合成し放送を行っていた。任意文を合成する際に，アクセントや読み仮名の推定に間違いが生じる可能性があるため，放送前に人手で修正を行っていた。そして，「多言語天気予報」は，2004 年 2 月 16 日から 2011 年 3 月 31 日まで，NHKデジタルラジオ実用化試験放送にて放送された。デジタルラジオのサブチャンネルを活かし，リスナーは日本語・英語・中国語・韓国語の 4 つの言語から選んで天気予報を聞くことができる。この他，番組の中での演出や効果として使われたものまで含めると，相当数あると考えられる。なお海外では，空港や駅での案内，高速道路での交通情報の放送などに音声合成を用いている例はいくつか見られるが，いわゆるテレビやラジオにおいて，合成音で放送した例は見当たらない。

上記の音声合成システムは，収録した音声を無音部分で接続して再生する方式(録音編集方式)がほとんどである[2-7]。録音編集方式で株式市況の音声合成システムを構築しようとすると，現在の最高値である数百万円までの数値を全て録音する必要があるが，膨大な時間とコストがかかり難しい。更に，録音編集方式では，録音した音声データの接続部分に適当な長さの無音が必要で，これがないと不自然な接続音になる。「株式市況」では，アナウンサーといえども容易ではないほどの早口で読み上げなければならないため，録音編集方式で必要となる長さの無音をはさむことはできない。

録音編集方式以外の音声合成方式も存在するが，肉声と同等の品質を実現できていない。HMM音声合成の自然性評価は文献[8]で行われている。ここでは，HMM音声合成システムを含む 14 種類の音声合成システムについて，同じ音声データベースを用いて同じ評価テキストから合成音を作成し，インターネットによる不特定評価者，音声の専門家，アメリカ英語を話す大学生による評価を行っている。評価は 5 段階で行われ，HMM音声合成は，インターネットによる不特定評価者では約 3.0，音声の専門家では約 3.1，アメリカ英語を話す大学生では約 3.4 という評価を受けている。その他の音声合成システムによる評価結果も公開されてはいるが，一番自然性が高いと評価を受けた合成音でも，インターネットでの不特定評価者では約 3.5，音声の専門家では約 3.7，アメリカ英語を話す大学生では約 3.7 という評価を受けている。(各音声合成システムの評価値は匿名で公開されているため，一番評価が高かった合成音が，どのような音声合成方式で実現されたかは不明である。) 自然音声が，ネットでの不特定評価者で約 4.5，音声の専門家で約 4.7，ネイティブスピーカーで約 4.4 と評価されたことを考慮すると，HMM音声合成方式では，肉声と同等の自然性は実現できていないことが分かる。

波形接続・波形編集の自然性評価は文献[9]で行われている。この文献では，2 つの評価実験が行われている。1 つ目の評価実験では，市販の 10 製品と提案法であるXIMERAで作成した合成音を 7 段階の評定尺度で評価を行った。その結果，波形接続・波形編集の他の 10 製品に対して，提案法であるXIMERAが統計的に有意に優れていることが分かった。2 つ目の評価実験では，XIMERAの音声データベースの大きさを 10 段階変化させた上で，5 段階の自然性評価を行っている。その結果，自然音声に対して約 4.8 という評価が得られているのに対し約 3.4 という評価が得られている。1 つ目の評価実験では，女性の発話者による 47 時間の音声データベースを用いており，2 つ目の評価実験では，男性の発話者による 63 時間の音声データベースを用いているという条件の違いはあるが，XIMERAも含め，従来の波形接続および波形編集では，肉声と同等の自然性は実現できていないことが分かる。また，これまでに筆者らは，ニュース番組の収録音声を音声波形データベースとして利用した波形接続型音声合成方式を開発し，良好な結果を得ている[11]。しかし，この方式でも放送に耐えうるような十分な自然性が得られるわけではない。

このため，新たに開発した株式市況音声合成システムでは，波形接続型音声合成方式を応用して，1 から 1 億未満の整数の数値音声について，桁単位の接続を想定し，調音結合(例えば「あいう」と発声したときの「い」には，前に発声した「あ」の口の形と舌の位置の影響と，後に発声する「う」の口の形と舌の位置の影響が表れている現象)を考慮して音声波形の接続を行った。これにより，音声の途中でのつなぎ合わせが可能になり，約 4000 個の数値を収録するだけで，1 億までの数値が合成可能になった。

また，この音声合成システムのもう一つの特徴として，放送用の音声合成システムとしては世界で初めて，合成音のしゃべる速度を変えることにより時間尺を調整したことが挙げられる。機械であれば株価データを受信した時点で，再生に必要な時間を計算できるので，音声と音声の間に均等に無音をはさむことで，ある程度の時間調整をすることができる。しかし，無音で調整できる時間には限りがあるし，「株式市況」で必要となる話速で元の音声を録音するのは，読み間違いや発声の安定性の面で難しい。そこで，話速変換を用いて，発声部分の長さを

変えることで，調整できる時間の範囲をより広くすることを行った。

3. 株式市況音声合成システムの開発の経緯と現状について

NHK技研では過去に音声合成の研究を行っていたが[10]，その後長い間研究を行っておらず，2001 年頃から音声合成の研究を再開した。再開後の研究では，大規模音声データベースを利用した波形接続型音声合成方式の開発を行い，「自然である」から「非常に気になる」の 5 段階で合成音の自然性を評価した結果，平均評定値で 4.01 を得る合成音を実現できるようになった[11]。しかし，平均評定値で 4.01 が得られたとしても，合成音の自然性は発話内容に大きく依存しており，放送に利用するレベルには到達していなかった。

そのような状況で，2003 年 6 月頃にNHKのアナウンス室から株式市況の自動化ができないか検討してほしいとの非公式の依頼があり，技研では同年の秋頃から株式市況の音声合成化の検討を開始した(なお，株式市況音声合成システムの実用化後に取り組んだ気象通報音声合成システム[12]についてもこの時に依頼があった)。研究は順調に進み，2004 年 6 月のNHK技研公開では，大規模音声データベースを利用した波形接続型音声合成システムの隣で，株式市況音声合成システムの展示も行った[13]。したがって，株式市況音声合成システムの骨格は，ほぼこの半年の間で完成することができた。

2005 年のNHK技研公開では，合成音の高品質化をさらに図り，運用イメージにより近づけた試作機の展示を行った。しかし，2004 年，2005 年と施設整備の予算申請を行ったところ，本当に安定して運用可能かどうか明らかでないとの理由から予算の申請は却下されてしまう。このため，地上デジタル音声放送実用化試験放送(通称デジタルラジオ)で実際に放送を出すことで，運用が可能であることを証明してみせることになった。

デジタルラジオは，当時受信機もほぼ販売されておらず，施設整備するための予算もついていなかったことから，技研の機材をそのまま持ち込み，株式市況のデータを自動的に取得するために必要な敷線工事・データ供給元のソフトウェア改修工事なども技研の主導で行った。これらの工事や運用に必要な改修は，主に 2005 年 10 月から 2006 年 4 月にかけて行い，2006 年 10 月から毎週金曜日に放送を行った[14]。ラジオ第 2 放送の株式市況は，当時の番組構成では，最初に 1 分程度概況を伝えて，それから各銘柄の株価と前日比を 44 分程度伝えるという構成であったが，デジタルラジオの放送にアナウンサーが配員される予定はなかったため，最初から各銘柄の株価と前日比を 45 分にわたり伝える完全自動放送の構成になった。

この放送は 2008 年 3 月まで行われたが，特に大きな不具合もなく，放送事故もなく，安定に運用することができた。運用がしにくい部分については，運用者に感想を聞いて改善点を列挙し順位付けをして，半年に 1 度システムの改修を行うことにより，改善を行っていった。

これにより，音声合成で放送を出すことが可能だと関係者に認知され，3 回目の予算申請でようやく認められることになり，ラジオ第 2 放送の株式市況の音声合成化が決定した。

決定してからも，関係する部局の担当者が何度も変わったり，システム整備の担当会社は入札で決定されるため，落札会社が必ずしも技研の技術を使うかどうかは決まっておらず，試作はしたものの別の方式で実用化される可能性があったりして，やきもきさせられることはたびたびあったが，最終的に技研で開発した技術が本運用機にも搭載されることになった。システムが整備された後，運用テストを何度か行ってリスナーの反響に問題がないことを確認した後，2010 年 3 月 29 日から本運用を開始した。

ラジオ第 2 放送で運用されたシステムについても，大きな事故を起こすことがなかったため，2014 年 3 月から 2 代目の株式市況音声合成システムがリリースされることになった。しかも，2 代目のシステムでは，気象通報の音声合成システムも搭載され，2 つの番組を音声合成で放送することが可能である。現在のところ，気象通報の音声合成システムはまだ放送に用いられていないが，株式市況の音声合成システムについては，2014 年 3 月 31 日から放送を行っている。

4. 数値音声合成

4. 1 数値音声合成の概要

数値音声合成の概要を図 1 に示す。

数値音声合成エンジンは数値が入力されると，4.2 節で述べる音声波形データベースの基本単位“クラスタリングされた前後の値を考慮した桁”に分割する。例えば，「1234 円」が入力された場合には，「一千(二百)」，「(千)二百(三十)」，「(百)三十四円」に分割する。ここで，「(千)二百(三十)」は，前が「千」で後ろが「三十」であるような「二百」を意味する。

次に，音声波形データベースの中には目的の数値を実

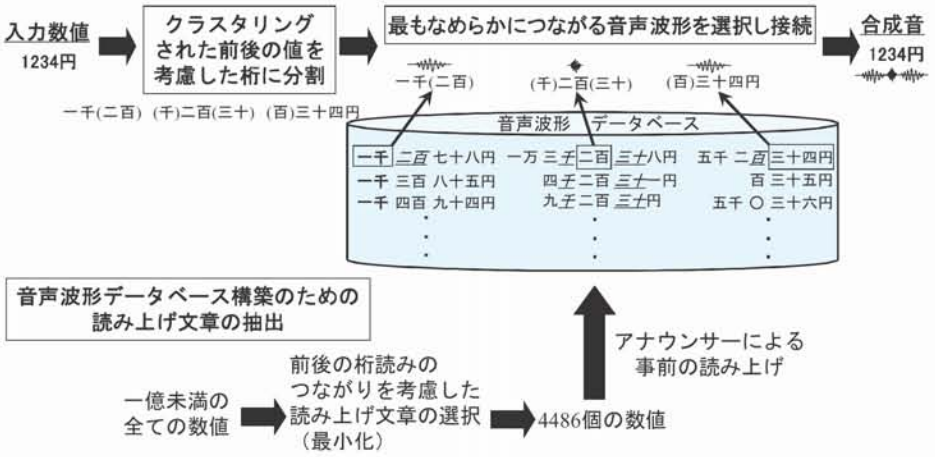

図１　数値音声合成の概要

現できる音声波形の組み合わせが複数存在するため，分割した桁を表す音声波形データのすべての組み合わせを探索し，隣り合う各音声波形データの音響的な特徴がなるべく類似する組み合わせを選択する。その結果，文全体で類似度の和が最大となる音声データの組み合わせを選択し，接続することで，合成音として出力する。

4．2　音声波形データベースの構築と基本単位

数値音声合成の音声波形データベースを構築する際，文献[11]のように目的の番組の放送音声を利用することが考えられる。しかし，株式市況の放送音声で使われる終値・前日比の数値には偏りがあり，放送音声を収録するだけでは，現在の放送を1年分収録したとしても，例えば1千万の音声は存在しないため，将来1千万を含む高額な終値・前日比が存在するようになった場合に音声を合成することができない。さらに，放送音声は，なるべく早口で読み上げようとするため発声が安定していない。このため，試作した株価音声合成システムの音声波形データベースは，数値の偏りがない文章をアナウンサーに別途読み上げてもらい構築した。この読み上げ文章の抽出法については 4.3 節で述べる。その際，発声が安定するように，株式市況の番組と比較してゆっくり読み上げるよう指示した。

また，目的は数値の合成であるため，“各桁”(下二桁，百，千，万下二桁，百万，千万)を音声波形データベースの基本単位とした。但し，単純に“各桁”とするのでは，無音をはさめないため合成音の自然性が劣化してしまう。このため，調音結合を考慮する必要があり，“前後の値を考慮した桁”にしたいが，この場合，基本単位の数が約４万と多くなる問題が生じる。したがって，下記の条件の桁は同じ桁とみなし，“クラスタリングされた前後の値を考慮した桁”として扱うことにする。

① 前の桁が十で終わる桁（十，二十，三十・・・九十)。これと同様に，前の桁が百で終わる桁（百，二百，三百・・・九百)，前の桁が千で終わる桁（一千，二千，三千・・・九千)，前の桁が万で終わる桁（一万，二万・・・九万，十万・・など）も同じ桁としてクラスタリングする。

② 後ろの桁が十で始まる桁（十一，十二，十三・・・十九)や二十で始まる桁(二十,二十一,二十二・・・・二十九)，同様に三十，四十，五十，六十，七十，八十，九十で始まる桁はそれぞれ同じ桁としてクラスタリングする。

上記のようにクラスタリングを行うことで，1 から 1 億未満の整数に含まれる基本単位の数を 5330 に削減することができる。

4．3　音声波形データベース構築のための読み上げ文章の抽出

音声波形データベースを構築するための読み上げテキストの作成方法について述べる。音声の収録を効率よく行うためには，読み上げテキストはできるかぎり少ないことが望ましい。しかし，本研究のタスクでは，各基本単位が音声波形データベースに1個以上存在しないと，その基本単位を含む音声が合成できなくなるため，読み上げテキストには必ず1個以上各基本単位を含ませる必要がある。

このため，読み上げテキストの中に各基本単位が一個以上含まれ，かつ大量のテキストから読み上げテキストがなるべく少なくなるように，以下のようなアルゴリズムで抽出を行った。

①あらかじめ定められている，1 文章に含まれる基本単位の最大数の初期値を 0 とする。

②大量のテキストから順に 1 文章ずつ選択し，その文章

に含まれている各基本単位の数をカウントする。
但し，すでに採用が決まった文章に含まれていた基本単位についてはカウントしない。

③カウントされた基本単位の数が，1 文章に含まれる基本単位の最大数以上の場合，この文章を採用する。
さらに，1 文章に含まれる基本単位の最大数をカウントされた基本単位の数に置き換える。

④②から③を大量のテキストに含まれる全ての文章について逐次繰り返し，採用する文章を増やしていく。

⑤④の操作を行った後でも，1 文章に含まれる基本単位の最大数が 0 であれば，採用された文章の基本単位は，大量のテキストの基本単位と一致するため終了する。
そうでなければ，採用した文章をそのまま保持し，①の初期化を行い②から③を再び繰り返し行う。

4．4　合成音評価実験

合成音の自然性を 5 段階で評価するため，作成した合成音に，3 つの市販の音声合成ソフトで作成した合成音(CA1，CA2，CA3 と表記)と，読み上げ音声を録音しただけの自然音声を加え品質評価実験を行った。3 つの市販の音声合成ソフトの音声合成方式は明らかではないが，波形接続方式では大規模な音声データベースが必要になるため，市販の音声合成ソフトとしては一般的ではなく，HMM音声合成方式も市販の音声合成ソフトにはほとんどないことから，波形編集方式である可能性が高い。

音声波形データベースは，1 から 1 億未満の整数を音声合成できるように，4.3 節の手法で抽出したテキストを読み上げることで構築した。抽出された読み上げテキストは 4486 個の整数である。読み上げたのは，株式市況の番組を担当している男性のアナウンサーである。録音は防音室で行った。

評価用数値は，音声データベースに含まれていない 40 個の株価と 40 個の前日比とした。1 つの手法で 80 個の合成音を作成したため，提案法・3 つの市販音声合成ソフト・自然音声の 5 つの手法を考えると，数値は全部で計 400 音声となる。

評価は，許容騒音レベルがNC-15 である防音室内でスピーカを用いて行った。聴取レベルはMCL(Most Comfortable Level)で，スピーカはDIATONEのDS-A3 を使用した。評定者は，音声の評価実験の経験のない 20 代の男性 10 名，女性 10 名の計 20 名である。各試行では，評価データをランダムな順序で提示し，評定者には自然性の違いを，5(自然である)，4(不自然な部分はあるが気にならない)，3(少し気になる)，2(気になる)，1(非常に気になる)の 5 段階で評価するよう指示した。自然性の評価では，合成音の品質評価に対するガイドライン[15]のように 7 段階の両極尺度で評価する手法もある。しかし，本論文では合成音の自然性のレベルを具体的に知りたかったため，文献[8]で行われているような 5 段階で評価することとした。評価に先立ち，音声波形データベース内の音声を 3 文章聞かせて，この程度の自然性の場合には，評価 5 の「自然である」と見なすようインストラクションを与えた。また，各音声の受聴は一回のみに限定した。なお，評価は，適度な時間間隔で休憩をはさみながら行った。

図 2 に各合成音のMean Opinion Score (MOS)と標準偏差を示す。自然音声の評価が 4.97 であるのに対し，合成音の平均評価値は 4.94 となった。一方，市販音声合成ソフトの評価は最も良いもので 3.44 となり，提案法で作成された合成音の自然性が十分に高いことが分かる。市販の音声合成ソフトよりも提案法による合成音の自然性が高いと評価された理由は，市販の音声合成ソフトが音響的な特徴量の目標値に音声データを変形することにより，合成音の自然性が低下してしまう一方，提案法では，そのような音声データの変形をする必要がない点にある。

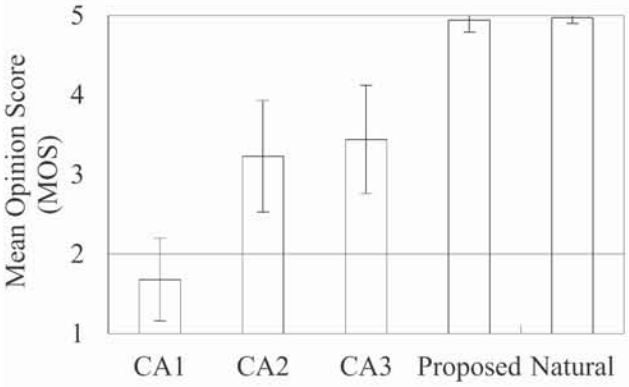

図 2　5 段階主観評価実験結果

5．むすび

ラジオ第 2 放送で運用している株価音声合成システムの音声合成部分について述べた。波形接続方式による音声合成システムを試作し，従来の録音編集方式では難しかった，収録した音声を有音部分で接続して合成音を作成することが可能になった。これにより，録音編集方式では実現できなかった番組の音声合成化が期待される。

また，話速変換技術を用いることにより，合成音のしゃべる速度を変えられるようになったため，調整できる時間尺の範囲が広がった。これにより，合成音の放送利用が一段と容易になったと思われる。

今後も引き続き実用的な音声処理技術の研究を推進す

ることで，社会に貢献していきたいと考えている。

参考文献

1) 日本放送協会,「20 世紀放送史」, 日本放送出版協会, pp. 31-32, 2001 年
2) 平岡 征男, 内山 久宜,「「早朝・深夜の天気予報」の番組制作 天気予報アナウンスコメント自動送出」, テレビジョン学会誌, Vol. 41, No. 8, pp. 742-743, 1987 年 8 月
3) 北浜, 今川, 三浦,「自動音声付天気予報システム」, 第 31 回民放技術報告会予稿集, pp. 94-95, 1994 年 11 月
4) 澤口 文夫,「天気情報アナウンス送出装置」, 映像情報メディア学会技術報告, Vol. 21, No. 53, pp. 25-30, 1997 年 9 月
5) 足立, 渡辺,「自動天気音声アナウンス送出システム」, 第 57 回「明日の放送と技術フォーラム」講演予稿集, pp. 27-28, 2004 年
6) 河上, 下野,「CG, 音声合成を用いた全自動道路交通情報システム」, 第 39 回民放技術報告会予稿集, pp. 148-149, 2002 年 11 月
7) 大田 雄二, 岩下 正信, 津浦 宏, 近藤 隆春, 大須賀 朋尚,「提供テロップ・アナウンスコメントの作成送出システムの概要」, テレビジョン学会技術報告, Vol. 17, No. 23, pp. 7-12, 1993 年 3 月
8) H. Zen, T. Toda and K. Tokuda, "The Nitech-NAIST HMM-based speech synthesis system for the Blizzard Challenge 2006", IEICE Trans. Inf. & Syst., Vol. E91-D, No. 6, pp. 1764-1773, 2008
9) 河井 恒, 戸田 智基, 山岸 順一, 平井 俊男, 倪 晋富, 西澤 信行, 津崎 実, 徳田 恵一,「大規模コーパスを用いた音声合成システム XIMERA」, 電子情報通信学会論文誌, Vol. J89-D-II, No. 12, pp. 2588-2698, 2006 年 12 月
10) 都木 徹, 梅田 哲,「ピッチ変更時のひずみをスペクトル領域で修正する音質変換方式とその品質の心理評価」, 電子情報通信学会論文誌, Vol. J73-A, No. 3, pp. 387-396, 1990 年 3 月
11) 世木 寛之, 田高 礼子, 清山 信正, 都木 徹,「ニュース番組の収録音声を利用した波形接続型音声合成システム」, 情報処理学会論文誌, Vol. 50, No. 2, pp. 575-586, 2009 年 2 月
12) H. Segi, R. Takou, N. Seiyama, T. Takagi, Y. Uematsu, H. Saito and S. Ozawa, "An Automatic Broadcast System for the Weather Report Program", IEEE Transactions on Broadcasting, Vol. 59, No. 3, pp. 548-555, 2013
13) H. Segi, R. Tako, N. Seiyama, and T. Takagi, "Development of a Prototype Data-Broadcast Receiver with a High-Quality Voice Synthesizer", IEEE Transactions on Consumer Electronics, Vol. 56, No. 1, pp. 169-174, 2010
14) 世木 寛之, 清山 信正, 田高 礼子, 都木 徹, 大出 訓史, 今井 篤, 西脇 正通, 小山 隆二,「高品質な株価音声合成装置の開発とデジタルラジオでの試験運用」, 映像情報メディア学会誌, Vol. 62, No. 1, pp. 69-76, 2008 年 1 月
15) 音声入出力方式標準化委員会,「音声合成システムの性能評価方法ガイドライン」, 電子情報技術産業協会, JEITA-IT-4001, 2003 年