



音声合成研究も協調と競争の時代に

—The Blizzard Challenge—*

徳田 恵一 (名古屋工業大学)**・アラン ブラック (カーネギーメロン大学)***

1. はじめに

新たな技術開発を行う際には、通常、各技術の比較評価を欠かすことはできない。ところが、任意のテキストから音声を作成するテキスト音声合成（以下、音声合成）に関して、各方式を比較・評価することは容易ではなかった。様々な要因が理由として挙げられるが、特に、それぞれの合成器の「声」の個性が異なる点が本質的な問題のひとつである。つまり、各機関の合成器は異なる音声データに基づいて構築され、従って、異なる「声」を生成する。このため、主観評価試験を行おうとしても、むしろ、声質に関する好みをアンケート調査するがごときものとなりかねず、各方式自体の性能や特性を適正に評価することを困難にしていた。この問題への取り組みのひとつとして、様々な音声合成方式を共通の音声データベースを用いて比較評価する Blizzard Challenge 2005 と題する評価会を開催した。本解説では、評価会の目的、設計、全体像等について述べながら、音声合成の比較評価に関する課題を「協調と競争」の観点から展望する。

2. 背景

音声合成と対になる音声認識の研究分野においては、共通のデータセットによる比較評価が有効に機能し、音声認識技術の進展に大きく貢献してきたことが知られている。特に、1990年代には、DARPA ワークショップにおいて、NIST 他が音声認識のテスト法を標準化した [1]。このような標準的な評価法と共通に利用可能なデータセット（学習用音声データベース及びテスト用音声デー

タ）の普及は、音声認識システムの性能の比較を容易とし、その結果、音声認識の中核技術の飛躍的な進展に寄与した。もちろん、単純な単語誤り率に基づいた評価に対する批判はあるものの、このような評価が音声認識技術を実用的なレベルに高めたという点については疑いの余地がない。

音声符号化の研究分野においても、セグメンタル SN 比等の客観評価尺度及び ACR (Absolute Category rating), DCR (Degradation Category Rating) 等の主観評価手法が確立されており [2], 各手法を比較的適正に評価することができる。音声符号化方式の標準化等において、このような評価手法は有効に機能したため、競争的な環境となり当該研究分野は大変な活況を呈した。

その一方、音声合成分野では、音声データのラベリング、音声単位の選択基準とそのためアルゴリズム、信号処理など、様々な要素技術を比較評価する必要があるが、分かり易い評価尺度がなかったこともあり、全体としてひとつの課題に取り組むコミュニティが形成されにくい状況が長く続いた。この頃には「音声合成は技術ではなく、むしろ芸術である」という言い方がされた。しかし、ここ 10 年ほどの間に普及してきたコーパス音声合成と総称される方式の台頭により、特殊技能を持った研究者が長い年月をかけて一つの「声」を持つ音声合成システムを構築するというスタイルから、音声データベースに基づいてシステムを自動構築するというスタイルへの移行が起こった。ところが、このようなシステムは、各研究機関ごとに特定のデータセットに対してチューニングされることとなるため、ラベリング、信号処理等の要素技術は、研究グループ内でのみ比較可能となっていた。また、各技術は、各機関が保有する音声データベースに最適化されるため、異なるタイプの音声データを用いた場合に有効であるかを確かめることは容易ではなかった。このようなデータセットとシ

* Speech synthesis research in a new age of cooperation and competition: The Blizzard Challenge.

** Keiichi Tokuda (Nagoya Institute of Technology, Nagoya, 466-8555)

*** Alan W. Black (Carnegie Mellon University)

システムの強い結びつきは、音声合成の様々な技術を適正に評価するための妨げとなっていた。つまり、各音声合成技術の性能や各システムから生成される音声の品質は、使用する音声データベースに大きく依存するため、異なるデータを用いて構築されたシステムにより、各技術を適切に比較することは困難となっていた。このため「構築された音声データベースそのものが音声合成システムである」、「新しい音声合成技術を開発するよりも、結果として良い合成音声を生成する音声提供話者を探した方が効果が高い」などと言われることすらあった。

このような観点から、音声合成技術のより良い比較を目的として、共通の音声データベースを用いた評価会 Blizzard Challenge 2005 が開催された [3]。以下、共通のデータセット、評価手法、参加者、受聴試験、受聴者など、音声合成技術の比較評価を行う際に問題となる事項を、主催者（及び参加者）として関わった立場から Blizzard Challenge 2005 での実際と関連付けながら順に述べる。

3. 共通のデータセット

Blizzard Challenge の最も重要な側面は、参加者が共有可能なデータセットを用いることである。音声データベースの違いを排除することにより、各システムから生成される合成音声をより適正に比較することが可能となる。ただし、このようなデータセットを用意する際には、様々な問題を考慮しなければならない。例えば、データのサイズ、形式、誰が収録し、公開するのか、独自の追加データの使用を許可するかどうか、などである。

音声合成コミュニティにおいて、共通のデータセットの重要性はこれまでも議論されてきたが、適切な諸条件を満たすものはなかなか現れなかった。音声認識研究のために設計された TIMIT 音素バランス 452 文は可能性のひとつであり、実際にエジンバラ大学音声技術研究センタは、特定話者が TIMIT 452 文を発声した音声データベースを公開したが [4]、通常の音声合成システム構築のためにはデータ量が不足する問題があった。ボストン大学の FM ラジオコーパス [5] はもうひとつの可能性であり、韻律関連の多くの実験に利用されたが、音素バランスがとられていないという問題があった。ATR 503 文 [6] も日本国内では広く

用いられたが、日本語であるということ自体と、海外からの入手が容易ではないという難点がある。その他にも、音声合成に関連した営利企業は、多くの場合、大規模で高品質な音声データベースを保有しているが、たとえ旧バージョンのデータであったとしても、データの内容や設計自体が知財として大きな価値があると考えられるため、それらを研究用として公開することは簡単ではない。

このような観点から、カーネギーメロン大学において、音声合成研究をより容易なものとする活動の一環として、約 1,200 文の音素バランス文からなる CMU ARCTIC データベースが公開された [7]。データは、グーテンベルグプロジェクト [8] による著作権の失効した多くの小説から選ばれた。まず、読み易く、適当な長さであり（5 単語から 15 単語）、すべての単語がカーネギーメロン大学が公開する発音辞書 CMUDICT [9] 内にある文が抽出された。これらから、ダイフオン（音素の二つ組み）の被覆率をできるだけ高くするように greedy に文を選ぶプロセスが 2 回繰り返され、それぞれ約 600 文ずつ、計約 1,200 文が選ばれた。

Blizzard Challenge 2005 では、2003 年に公開された 4 話者分の CMU ARCTIC データベースから、アメリカ英語の男性話者 1 名、女性話者 1 名の計 2 話者分の音声データが選ばれ、学習用データとして用いられた。また、各参加チームがいかに早くシステムを構築することができるかを試すために、合成音声提出 1 週間前に、新たな 2 話者（男性話者、女性話者それぞれ 1 名）が公開された。

なお、CMU ARCTIC データベースは今回の目的に対し適切な開始点と考えられるが、参加システムによっては、より大きなデータを前提としたものもあることに注意する必要がある。特に、単位接続型の音声合成システムでは、より大きなデータを用いることの利点を活かしたものとなっている。また、発声内容も、CMU ARCTIC が小説からなっているのに対し、様々なドメインの文からなっていることが多い。

4. 評価手法

これまでも、音声合成の評価に関して様々な検討がなされているが、各方式や技術自体を評価するというよりは、むしろユーザの立場から音声合成製品の品質や特性を評価しようとするもので

あった [10]。あるいは、ウェブによるオンラインデモなどが普及する以前には、同じテスト文で様々なシステムを比較することを主目的とするものもあった [11]。

Blizzard Challenge の目標は、同じ学習用音声データに基づいて構築されたシステムを比較することにより、様々な条件の下で、より良い音声合成技術を見つけることであるが、どのような意味で「良い」かは一意ではない。これまでの音声合成の評価では、読みや発音の精度、人名や記号等の読み上げ精度、自然性、明瞭度等、様々な観点からの評価尺度が検討されているが、Blizzard Challenge 2005 では、最も単純なものが幾つか選ばれた。

まず、五つのジャンルが設定され、それぞれに対するテキストが選定された。ここでは、数字や記号の読み付与等、テキスト解析における諸問題の議論を主目的とはしないとの趣旨から、数字や記号をあまり含まない比較的単純なテキストが選ばれた。なお、参加機関の間でのテキスト解析器の共通化は行われなかった。これは、用いる音素セットが異なるなど、参加機関ごとにテキスト解析器に要求される仕様や条件が大きく異なるためである。ただし、独自のテキスト解析器を持たない参加機関は、Festival 音声合成システム [12] のそれを用いることができ、実際に半数以上の参加チームが Festival のテキスト解析器を用いた。

初めの三つのジャンルは、ACR 試験により得られる MOS (mean opinion score) による評価の対象である。つまり、受聴者は、各合成音声サンプルを 1~5 点で評価することとなる。残りの 2 ジャンルは、受聴者自身が聞き取ったものをタイプする書き取り試験のためのものであり、書き取り結果の単語誤り率が音声認識における性能評価と同様に算出される。従って、主として明瞭度の評価を行うことになる。

以下、五つのジャンルの概要と例文を示す。

【小説】 CMU ARCTIC 構築に用いられた幾つかの小説から選ばれたテキスト。ただし、CMU ARCTIC で用いられた文は除く。ドメイン内タスクに対応する。

- *Joe Garland lives like a good fellow.*
- *But we made no collections of eggs.*

【ニュース】 標準的なニューステキスト。

- *The two countries agreed to resolve any*

conflict through diplomacy and avoid the use of force, the agency Interfax said.

【会話】 音声対話システムで用いられるような発声。ただし、ユーザ側の発声に対応するもの。従って、通常の音声合成におけるテキストとはかなり異なるものとなる。

- *Okay I would like to go to Miami, Florida.*
- *Yeah I guess it will and something downtown please.*

【音韻的に混同し易い文】 MRT (Modified Rhyme Test) [13] と同様、音韻的に混同し易い単語の聞き取りを目的とする。ただし、オリジナルの MRT のように孤立単語ではなく、混同し易い単語を定型文の中に挿入したものである。これは、単位選択等の方式では、孤立単語発声の場合とその単語が文内にあるときで、異なる音声単位が選択されうるからである。

- *Now we will say cold again.*
- *Now we will say pace again.*

【無意味文】 文献 [14] に倣い、簡単な文法テンプレート (定冠詞+形容詞+名詞+動詞+定冠詞+形容詞+名詞) に従い、出現確率が中程度の単語がランダムに生成された。得られる文は、人間が発話したとしても、理解することも記憶することも容易ではないものとなる。

- *The unsure steaks overcame the zippy rudder.*
- *The dank geniuses woke the humane emptiness.*

それぞれのジャンルに 50 文ずつ、計 250 文が、テスト文として選定あるいは生成された。なお、実際の文選定・生成は、評価会の参加者でもある主催者からの独立性を高めるため、UIUC の Richard Sproat 教授により行われた。

5. 参加者

当初、5 大学、3 企業、計 8 機関の参加が見込まれたが、うち 1 企業は直前に買収されたため、参加取り止めとなった。また、もう 1 企業は結果の公表に関する心配から参加を取りやめた。このため、最終的に 5 大学、1 企業、計 6 機関が参加することとなった。参加機関の所在地は、英国、米国、日本に分布し、企業は米国の大企業であった。本

評価会は、あくまで、様々な技術を持ち寄り、評価することによって得られる知見を共有することが本来の目的であり、各社の音声合成製品に優劣をつけようとするものではないが、にも関わらず、このような評価会に参加することは、営利企業にとって大きなリスクと見なされることに注意を払う必要がある。

6. チャレンジ

評価会のスケジュールは以下のとおりであったが、初めての試みでもあり、必ずしも予定どおりに進まない部分もあった。

2004年

12月 2日 Blizzard Challenge 2005 告知

12月 15日 参加者登録

2005年

1月 15日 新しい音声データベースの公開

1月 15日 テスト文の公開

1月 22日 合成音声提出

2月 1日 受聴試験による評価開始

2月 28日 受聴試験による評価終了

3月 15日 参加者へ評価結果を配布

4月 8日 スペシャルセッション論文締め切り

9月 5日 Interspeech2005 スペシャルセッション

2005年1月15日に参加者にテスト文が公開された。また、既に公開済みであった2話者分の音声データに加え、更に2話者分の音声データがテスト文と共に公開された。テスト文250文を合成するために、各チームには1週間の期間が与えられた。参加者がシステムをテスト文にチューニングする余地があったが、特に対策は施されなかった。なお、6チームのうち、1チームは1週間のうちに新しい音声データベースによる「声」を構築することができなかったが、以下の議論においては、音声サンプル数等の説明の都合上、すべての「声」が構築できたと仮定することにする。

六つの参加チーム以外の仮想参加チームによる合成音声として、音声データ提供話者が実際にテスト文を読み上げた音声の評価対象に加えられた。これにより、現在の音声合成システムが自然音声からどれほど離れているかを見積もることができる。

7. 受聴試験

受聴試験は、受聴者の数と音声サンプル数の多さから、ウェブベースで行われた。もちろん、ウェブにより行われる受聴試験には、様々な欠点がある。ひとつは、受聴者の受聴環境が、それぞれ大きく異なることである。受聴者は静かな部屋での受聴を求められたが、実際にそのような条件下で受聴が行われたことを確認するのはほとんど不可能である。また、ネットワークの状況によっては、音声波形ファイルのダウンロードに時間を要するのみならず、音の途切れ等の発生が考えられ、実際にそのようなことが起こったとの報告があった。更には、受聴者が試験の途中で興味を失い、途中で試験を中断したり、でたらめなスコアを入力したりすることも考えられる。

これらの欠点にも関わらず、ウェブによる評価試験には大きな利点がある。通常、数百名の受聴者を集めることは容易ではないため、できるだけ簡単により多くの受聴者を集める工夫が必要となる。実際、Blizzard Challenge 2005では、多くの国々から多数の受聴者を募ることができた。また、ウェブによる受聴試験の実現可能性についての議論は以前よりあり、その実証を行うこともなる。これらの観点から、ウェブによる評価試験には、予想される精度低下を補って余りある利点があると考えられる。

受聴試験は、Blizzard Challenge ホームページ [3] を訪れることにより行われた。試験は五つからなり、それぞれがひとつのジャンルに対応する。各参加チームは、各ジャンルに対し50の発声を4種の「声」（男女×新旧）により合成したが、受聴者ごとの試験の規模が大きくなり過ぎないようにするため、受聴者ごとにひとつの「声」と20テスト文がランダムに選ばれ、対応する合成音声各受聴者に提示された。従って、受聴者は、各試験で、7システム分（6参加システムと自然音声による仮想参加システム） $140 = 20 \times 7$ の音声サンプルを受聴することになる。それぞれのサンプルは、一度だけ提示され、スコアを入力しなければ次の音声サンプルに進むことができないように設定された。試験は、一時期に行うよう勧められたが、途中で一時中断、後日、再開できるように工夫された。各受聴者は、五つの試験に対応する

計 700 サンプル (140 サンプル × 5 ジャンル) を評価することになり、このために通常 30~45 分程度を要した。

8. 受 聴 者

誰が受聴者として音声合成システムを評価すべきかという点も未解決の問題のひとつである。当該分野の専門家は、詳細な知識を持ち、イントネーション、音声単位の接続誤差、音素誤り、あるいは、信号処理に起因する異音、等を聞き分けることができる。一方、現実のユーザは、誰の声に聞こえるか、自分の好みの声質かどうか、等に関心があり、不自然に聞こえる合成音声の原因を同定することはできない。従って、より安定した結果を得るためには、多くの異なるタイプの受聴者が必要と考えられる。

このため、以下に示す三つの異なる受聴者のグループが設定された。

【音声専門家】 各参加機関は、10名の音声専門家を受聴試験に参加させることを義務付けられた。音声合成に関して最も専門的な知識を持ち、各機関が自身のシステムを評価する際に照会すると予想されるグループである。各受聴者は、自身が所属する機関のシステムを含めて評価した。

【ボランティア】 評価システムは、受聴試験に誰でも参加可能となるようインターネット上で公開され、音声関連メーリングリストやウェブページにより受聴者の募集が行われた。全くの一般の人々から、音声専門家、更には音声合成に携わる専門家に至るまで、様々な人々が含まれる。

【アメリカ学部学生】 3番目のグループは、アメリカ国内において、アメリカ英語を話す学部学生のグループである。他の2グループと異なり、このグループの受聴者には、若干の賃金が支払われた。これは退屈な仕事に対して一貫した結果を得るためである。

これら、三つのグループはそれぞれ、全く異なる目的意識を持って、試験に参加している。グループ間の違い、また、グループ内での一貫性に関して、調査することは大変興味深いことである。

第1のグループは最も熱心に試験に取り組み、試験を実直に完了させた。それに対して、第2のグループは、最後まで試験を終了させるものもいれば、途中で放棄してしまうものもあり、ばらつ

きが大きかった。第3のグループは、予想に反し、試験に参加させること自体が、それほど簡単ではなかった。五つの試験の完了に対する報酬として、5米ドルのamazon.comギフトを、更に、同じ五つの試験をもう一度行ったものに対しては、追加の10米ドルのギフトが提供された。これは、学部学生を同様の実験に参加させるために支払われる額とほぼ同じであるが、目標であった延べ100名の学部学生の参加を得るにはやや不足していたようである。

9. 予測しなかった問題

新しい試みには必ず予期しない出来事が起こるものである。今回起こった問題のひとつは、与えられた音声データ以外のデータをどれほど用いることが許されるかという点であった。システムによっては、韻律モデルが特定の話者の音声データに基づいて構築され、異なる「声」を構築する際にも固定のままであることがある。このような場合には、与えられたデータ以外を用いていることになるため、このような「割り増し」データを禁止することは必ずしも適切ではないと考えられる。

実際に想定外だったのは、声質変換技術を利用したシステムである。つまり、大量のよく整備された音声データに基づいて構築された高品質な音声合成システムに対して、評価会用に与えられた音声データを学習データとして声質変換を適用するというものである。このような参加システムがあることが明らかになったのは、ACRによる受聴試験を設計した後であったため、声質や発話スタイルの類似性を評価できないという理由から、このようなエントリーは今回認められないこととなった。音声提供話者との類似性を考慮しない今回の評価方式では、与えられた音声データを全く利用せず、評価会とは別につくられた大企業の最高品質のシステムをエントリーしたとしても高評価を得ることになってしまうという問題がある。

いずれにせよ、声質変換は音声合成システム構築のための有効な手段であることは確かであり、声質変換に基づいたシステムを有効なエントリーとするには、何等かの形の類似度試験を導入する必要がある。

10. 結 果

得られた多くの様々な知見は, Interspeech2005 [15] のスペシャルセッションにおいて, 8 件の論文として発表された。5 章の議論から, 参加機関の名前とそれぞれのランキングとの関係は明らかにされず, 最終結果は, 各機関を表す一文字のアルファベットによって公表された。各機関は, 自身に対応するアルファベットのみを知ることができる。ただし, 各発表論文の内容を相互に見比べることにより, それぞれのアルファベットと参加機関との対応関係を推定できる可能性がある。

評価会を通し, 各合成音声サンプルが延べ 10 名以上の受聴者によりスコア付けされた, 50 発声 × 5 ジャンル × 7 システム × 4 話者分, 計 7,000 の合成音声サンプルが得られた。これらのデータは, 参加機関のみならず, 合成音声の自動評価に関する研究等にも大いに役立つものと考えられるため, 今後, 一般に公開することが検討されている。

11. 将来のチャレンジへの提言

このようなチャレンジが例年行事となることが期待されるが, 音声データベースを固定した場合には, 音声データベースに依存した画一的な技術開発となりかねないため, 様々なタイプの音声データベースを用いる必要があると考えられる。評価手法も様々な形で拡張していくことが必要であろう。このような観点から, 英語以外の言語に対しても, 同様の試みが行われる予定である。また, 予告なし言語 (予告なく知らされた言語に対し, 音声合成システムを短時間で構築する), 声質変換, 制限フットプリント (組み込み用途を想定し, 厳しい計算機資源の制約の下で動作可能な音声合成器を構築する), 制限音声データ (100~数百発声程度の極度に少ない音声データによりシステムを構築する) 等, 様々な提案が既になされており, 今後の興味は尽きない。

ただし, 頻繁なルールの変更は, 参加者を混乱させ, 前回の結果からのフィードバックを困難にするとも考えられるため, 少なくとも 3 回程度は, 類似したルールによる評価を繰り返す必要があるだろう。

Blizzard Challenge 2006 に向けて考えられる変更点は以下のようなものである。

【より大きなデータベース】 より大きな音声デー

タベースを前提としたシステムを持つ参加機関は, 自身のシステムに適したサイズのデータベースを利用したいと考えるであろう。幸い Blizzard Challenge 2006 に向けて, ATR 音声言語コミュニケーション研究所より, 単一話者により発声された 5 時間にわたる高品質音声データベース (アメリカ英語) が評価会用に提供されることになり, 準備が進んでいる。

【類似度試験】 コーパス音声合成の究極的な目標は, あらゆるテキストに対し, あらゆる意味で音声提供話者の発声と区別のつかない音声を生成することであろう。この観点から, 音声提供話者との類似性を評価するための DCR 試験を導入する必要がある。これにより, 声質変換に基づいたシステムについても, 適正に評価を行うことができる。ただし, 評価試験のコストは増加せざるを得ないことに注意する必要がある。

12. む す び

Blizzard Challenge 2005 は, 共通のデータセットから音声合成システムの「声」を構築する大変有益な演習となった。また, 全く異なる研究グループが共通のゴールを目指すことを可能としたという点で大変意義深いものとなった。我々自身も, 明確な目的が設定されることによる研究グループ内の意欲向上, メンバ間のコミュニケーションの拡大等, 予期した以上の多大な良い効果を実体験した。このような大規模な評価実験は, 独立した複数の機関の協力なくしてはありえず, 結果として多くの貴重な知見を共有することができた。音声合成研究も, 音声認識などの分野と同様, 共通の研究基盤に基づいた協調と競争の時代に入ったと言えるのではなかろうか。

Blizzard Challenge 2005 は, 今後の一連の評価会の始まりと考えられ, 既に, Blizzard Challenge 2006 [16] への活動が開始されている。前回以上に多くの参加申し込みがあり, 本稿が読者の手に届く頃には, 第 2 回 Blizzard Challenge の結果が出揃っているはずである。なお, 結果は Interspeech2006 [17] のサテライトワークショップとして発表される予定である。

謝 辞

The Blizzard Challenge 2005/2006 開催に際し, 音声データ提供, テスト文設計, ウェブ評価

システムの構築, 受聴者としての米国学部学生の手配, 受聴試験への参加, チャレンジへの参加ほか, 様々な形でご協力いただいた多くの方々に深く感謝したい。

文 献

- [1] <http://www.nist.gov/speech/publications/>
- [2] W.B. Kleijn and K.K. Paliwal, *Speech Synthesis and Coding* (Elsevier, Amsterdam, 1995).
- [3] <http://festvox.org/blizzard/blizzard2005.html>
- [4] http://www.festvox.org/dbs/dbs_kdt.html
- [5] M. Ostendorf, P.J. Price and S. Shattuck-Hufnagel, "The Boston University Radio News Corpus," *Boston University Technical Report*, ECS-95-001, March (1995).
- [6] Y. Sagisaka, M. Abe, T. Umeda, S. Katagiri, K. Takeda and H. Kuwabara, "Large-scale Japanese speech database," *Proc. ICSLP 90*, pp.1089-1092 (1990).
- [7] http://festvox.org/cmu_arctic/
- [8] <http://promo.net/pg/>
- [9] <http://www.speech.cs.cmu.edu/cgi-bin/cmudict/>
- [10] 日本電子工業振興協会, 音声合成システム性能評価方法のガイドライン: JEIDA-G-24-2000: JEIDA ガイドライン (日本電子工業振興協会, 東京, 2000).
- [11] J.P.H. van Santen, L.C.W. Pols, M. Abe, D. Kahn, E. Keller and J. Vonwiller, "Report on the Third ESCA TTS Workshop Evaluation Procedure," *Proc. 3rd ESCA Workshop on Speech Synthesis*, pp.329-332 (1998).
- [12] <http://www.cstr.ed.ac.uk/projects/festival/>
- [13] A.S. House, C.E. Williams, M.H.L. Hecker and K.D. Kryter, "Psychoacoustic speech tests: A modified rhyme test," *Tech. Doc. Rep. ESD-TDR-63-403*, U.S. Air Force Systems Command, Hanscom Field, Electronics Systems Division (1963).
- [14] C. Benoit, M. Grice and V. Hazan, "The SUS test: A method for the assessment of text-to-speech synthesis intelligibility using Semantically Unpredictable Sentences," *Speech Commun.*, 18, 381-392 (1996).
- [15] <http://www.interspeech2005.org/>
- [16] <http://festvox.org/blizzard/>
- [17] <http://www.interspeech2006.org/>
-