2005－SLP－55　(9)
2005／2／4

音声認識応用に関する学会試行標準

○西本 卓也 †1　○西村 雅史 †2　赤堀 一郎 †3　石川　泰 †4
磯谷 亮輔 †5　伊藤 克亘 †6　大淵 康成 †7　金澤 博史 †8
國枝 伸行 †9　外山 聡一 †10　新田 恒雄 †11

†1 東京大学　†2 日本アイ・ビー・エム (株)　†3 (株) デンソー　†4 三菱電機 (株)
†5 日本電気 (株)　†6 名古屋大学　†7 (株) 日立製作所　†8 (株) 東芝
†9 松下電器産業 (株)　†10 パイオニア (株)　†11 豊橋技術科学大学

†1 東京大学大学院 情報理工学系研究科, 〒 113-8656 東京都文京区本郷 7–3–1, nishi@hil.t.u-tokyo.ac.jp

あらまし　本報告では，情報処理学会試行標準委員会の WG4 小委員会（音声言語インタフェース）が策定した，(A) ディクテーションに用いる基本記号に対応する読みの試行標準，および (B) カーナビ用音声入力の性能評価ガイドラインの試行標準検討案，の内容を紹介する．

キーワード　音声言語処理，音声認識，ディクテーション，カーナビゲーション，標準化

IPSJ Trial Standard for Speech Recognition Applications

Takuya NISHIMOTO †1　Masafumi NISHIMURA †2　Ichiro AKAHORI †3　Yasushi ISHIKAWA †4
Ryosuke ISOTANI †5　Katunobu ITOU †6　Yasunari OBUCHI †7　Hiroshi KANAZAWA †8
Nobuyuki KUNIEDA †9　Souichi TOYAMA †10　Tsuneo NITTA †11

†1 The Univ. of Tokyo　†2 IBM Japan Ltd.　†3 DENSO CORPORATION　†4 Mitsubishi Electric Corp.
†5 NEC Corp.　†6 Nagoya Univ.　†7 Hitachi Ltd.　†8 TOSHIBA Corp.
†9 Matsushita Electric Industrial Co., Ltd.　†10 Pioneer Corp.　†11 Toyohashi Univ. of Tech.

†1 7–3–1, Hongo, Bunkyo-ku, Tokyo 113-8656 Japan

Abstract　This paper describes the activity of IPSJ/ITSCJ Trial Standard WG4 that has the objectives of standardization for spoken language interface. In this report, we introduce : (A) trial standard of sounds-like spellings of commonly used symbols in Japanese dictation systems, and (B) working draft of guideline for evaluating speech input performance of car navigation systems.

Keywords　Spoken language processing, Speech recognition, Dictation, Car navigation, Standardization

1　はじめに

平成 14 年に情報処理学会に発足した学会試行標準専門委員会の下で，WG4 小委員会（音声言語処理インタフェース）が活動を行っている．学会試行標準の目的は次の 2 点にある [1].

- 国際標準を成立させるには長時間を要するので，準備段階のものを学会として標準化する．
- 国際標準の基礎となるデータで，国際標準として制定が難しかったり，国際標準になじまないものを学会標準とする．

WG4 小委員会の活動目的は，音声入出力を利用する際におこる様々な問題を，標準化の観点から解決することである．これまでに，音声認識の対象単語に用いられる読み表記，および音声認識・合成に関する用語の検討を行った [2, 3]．本稿では音声認識応用に関する新たな活動として，ディクテーションソフトウェアに関する試行標準と，カーナ

ビゲーションシステム（カーナビ）に関する試行標準案について報告する．

2 ディクテーションに用いる基本記号に対応する読み

2.1 背景と目的

ディクテーションに用いる基本記号に対応する読みには規準がないため，製品での対応は各社まちまちである．そこで，2003 年 7 月以降 WG4 小委員会は「音声認識応用ソフトウエアに係わるコマンド集合」の中で,「ディクテーションに用いる基本記号に対応する読み」を早期に学会試行標準とする作業を開始し，主としてキーボードから入力可能な基本記号に対応する読みを規定した．この試行標準は 2004 年 11 月に公開された [4].

本試行標準に準拠したシステムが普及することにより，ユーザがどの日本語ディクテーションシステムを使用していても同じ読みで同じ文を入力できることが期待される．今後，業界団体等において，本試行標準をベースに規格化へ向けた議論が高まることを望む.

2.2 試行標準の詳細

本試行標準では，ディクテーションに用いる基本記号に対応する読み (Sounds-like Spellings of Commonly Used Symbols in Japanese Dictation Systems) を,「ディクテーションシステムにおいて文入力中に基本記号をどのように読むかを規定するもの」と定義する．読みの記述には「音声認識のための読み表記」[3] における 「読み表記」 (Sounds-like Symbols) を使用する．

主にディクテーションシステムによって日本語の文を入力する場合に用いられる，基本記号に対応する読みを表 1 に示す．この試行標準に適合するディクテーションシステムは，表 1 に示す「基本記号」に対して，表 1 の「読み」を認識するように設計されなければならないが，その他の「読み」を認識してもよい．

この試行標準に適合するディクテーションシステムが表 1 の読みを正しく認識した場合に，入力される基本記号が，複数の文字コードに対応している場合（例えば，2 バイト文字と 1 バイト文字）は，アプリケーションに依存して区別が行われるものとし，本試行標準では特に規定しない.

表 1: ディクテーションに用いる基本記号に対応する読み

基本記号	読み	
,	こんま	
.	ぴりおど	どっと
:	ころん	
;	せみころん	
?	くえっしょんまーく	
!	びっくりまーく	
^	はっと	
～	から	
~	ちるだ	
_	あんだーばー	
ー	まいなす	はいふん
／	すらっしゅ	
＼	ばっくすらっしゅ	
｜	たてぼう	
＋	ぷらす	
＝	いこーる	
＜	しょうなり	
＞	だいなり	
¥	えん	
$	どる	
%	ぱーせんと	
#	しゃーぷ	
&	あんぱさんど	あんど
＊	あすたりすく	
@	あっとまーく	
(空白)	すぺーす	
’	くおーてーしょん	
”	だぶるくおーてーしょん	
（	かっこ	
）	かっことじる	
［	だいかっこ	
］	だいかっことじる	
「	かぎかっこ	
」	かぎかっことじる	
、	てん	
。	まる	
・	なかてん	

2.3 主な検討内容

本試行標準の作成においては，基本記号の定義と，標準的な読みの選定についての検討が行われた.

まず，基本記号の範囲としては，一般的なパーソナルコンピュータの日本語キーボードによって入力可能なもののうち，英数字，漢字，かたかな，ひらがななどの文字を除いたものを目安として選択し,「改行」などの操作の名称は含まないこととした.

「ー」と「-」（全角および半角のマイナス）のように，複数の記号の視覚的形状が同じである場合は，これらをまとめて 1 種類の基本記号として扱い，これらはアプリケーションに依存して区別が行われるものとした.

次に，標準的な読みの選定にあたっては，すでに市販されている複数の日本語ディクテーションシステムの仕様を検討した[1]．特に，エンドユーザが利用できることを念頭において， 覚えやすく読みやすい読みであること，関連する基本記号の読み同士に一貫性があること，1 つの読みが 1 つの基本記号にのみ対応すること，既存のシステムでできるだけ広く使用されていること，などを考慮した．

既存のシステムにおいては，特に「,」や「＊」などに関しては多くのバリエーションがあった．そこで，特に外来語を語源とする読みについては，その語源となる単語に置き換えて広く用いられている単語を選択し， さらにその単語に対して広く用いられている読みを選択した．例えば「＊」の読みとして「あすたりすく」「あすてりすく」「すたー」などが既存のシステムで採用されているが，まずこれらを ”asterisk” ”star” などの語源別に集計し，最も広く採用されている”asterisk” を選択した．次に ”asterisk” の読みとして広く採用されている 「あすたりすく」を選択した．

あるアプリケーションにおいて，この試行標準が定める 1 つの読みが入力された場合に，実際にどの基本記号が入力されるかは，エンドユーザの便宜を考慮して定めるべきである．例えば，ワードプロセッサの句読点モードを変えることで,「てん」「まる」を「、」「。」として入力したり,「,」「.」として入力することができるが，この選択はユーザの嗜好に属する．

今後は次の事項について検討する予定である．

- 視覚障害者用スクリーンリーダにおけるキーボード操作の音声読み上げ方法
- アプリケーションの操作に関するコマンド集合（例：「改行」「新しい段落」など）
- かな文字を 1 文字ずつ区切って音声入力する場合のコマンド集合（例：「朝日のあ」「いろはのい」など）

3 カーナビ用音声入力の性能評価ガイドライン

3.1 背景と目的

WG4 小委員会は 2004 年 1 月に，カーナビに使用される日本語音声入力の性能評価に係わるガイドラインの作成に着手した．現在も検討は続いているが，本稿では，現時点での学会試行標準（案）について述べる．

現在，多くのカーナビゲーションシステムが音声認識機能を有し，音声は有効な入力手段の一つとなっているが,“ 多くのユーザーが日常的に使っている ”あるいは“ 性能に満足している ”という状況ではないこともよく知られた事実である．

カーナビに使用する音声入力ユニットに対しては，その性能を評価する際の共通ガイドラインが存在しないため，各社で比較不能な評価が行なわれている．一方で，ユーザに対しては，音声入力がどの程度使える機能かといった判断材料が示されていないため，カーナビ購入時の比較検討項目にこの機能を入れることができない．技術は日々進歩しているにもかかわらず，正しい使用方法での評価が行われなかったために音声認識は使えないというイメージだけが定着している恐れもある．このような状況は健全な技術競争を生み出さず，カーナビメーカーとその購入者だけでなく，音声認識技術の普及発展にとっても残念な状態だと言える．

これまでディクテーションソフトなどの音声認識製品では，雑誌社などの第三者が独自の基準で比較を行い，ユーザ層に対し多くの判断材料を提供していた．適切な評価基準がなかったため，その比較方法には問題があった面もあるが，性能評価が容易な対象であったこともあり，おおむねユーザの立場からの適切な評価がなされていたと考えられる．一方，カーナビの認識装置は使用状況が複雑多岐にわたる上，スイッチの操作方法，コマンド名称なども各社まちまちで，統一的な認識性能評価は困難である．また，雑音下のハンズフリー大語彙連続音声認識といった非常に難しいタスクも対象に含まれるため，環境の影響を大変受けやすく，正しい手順に従った評価をすることが特に重要である．

音声認識は他の入力手段（たとえばタッチパネル）ほど外乱に対してロバストではないが，一方で，目的地を選ぶ場合のように，階層をたどることなく，すぐに入力が完了するといった利点を持ち合わせている．ともすると，第三者による評価は，このような音声認識の利点を活用するという立場よりも，他の入力手段と比べた場合の音声認識の欠点の調査に陥りがちであるので，評価の手順を明確に定めたガイドラインが必要であると考える．共通のガイドラインが試行標準として提供されると，各認識システムが本来持つ能力の比較が可能になる．このことは，最終的には音声入力ユニット開発各社およびユーザにとっても有益と考えられる．

[1]調査対象とした製品は IBM ViaVoice, TOSHIBA LaLaVoice, NEC SmartVoice, ScanSoft Dragon Speech, Microsoft Japanese ASR である．

3.2 ガイドラインの概要

3.2.1 評価対象機器

本試行標準（案）は，音声入力機能の付いたカーナビゲーションシステムもしくは車載情報機器を評価対象機器とする．

3.2.2 評価の対象となる機能

音声入力の評価は，現状の音声入力付カーナビにおいて，音声入力の使用頻度が高く，かつ他の入力手段と比較して有用と考えられている POI（Point of Interests: 関心地点）の認識率のみで行う[2]．詳細住所入力や電話番号入力なども音声入力の有効なアプリケーションではあるが，カーナビによってポーズを入れる位置が異なるなどの事情があり，現状では認識性能の対等な比較は困難である．なお，システムの“ ユーザビリティ ”評価は本ガイドラインには含まない．

3.2.3 性能評価方法

ここでは簡易評価と推奨評価の 2 つの評価方法を定義する．

簡易評価方法 雑誌社など第三者が複数のカーナビを比較評価することを目的に策定するものである．性能評価の際に守るべき一定の基準を示すことで，誤った使い方によって，誤った性能評価が行われることを避けることができる．また，評価者の負担がなるべく小さくなるように考慮することで，本基準が広く利用されることを目指す．たとえば，評価者自身が被験者になることも許容する．

推奨評価方法 大規模かつ詳細な実験条件を定義することで，認識率の絶対値に意味を持たせる．本来，第三者や評価機関による評価が最も望ましいが，カーナビメーカー自体がこの評価基準で認識率を測定し，それをパンフレットに記載するなどの利用を想定している．

本評価方法に従った実験においても，認識対象語彙に関する条件はシステムのユーザビリティと関係する項目であって，これを各社同一条件にすることは不可能である．メーカーがこの条件下での対象語彙サイズを実験結果に併記するなどの対応が必要である．これについては後述する．

[2]POI の例を挙げると「東京駅」「東京タワー」「上野動物園」「国立西洋美術館」「羽田空港」「塩原グリーンビレッジ」「海老名サービスエリア」などがある．現在値に応じて認識対象となる POI を制約する製品が多い．

3.3 ガイドラインの詳細

3.3.1 簡易評価方法

発話者（被験者）

発話者は日本語を母国語とする男女各 1 名（以上）とする．実験の評価者自身を発話者に含めてもよい．

発話リスト

添付（予定）の POI リスト（100 箇所）などを参考にして，各話者 50 箇所（以上）を選んで認識実験を行う．なお，後述するように自車の現在地を“ 新宿，東京都庁 ”として実験を行うため，添付（予定）の POI リストには多くのカーナビで入力可能な東京周辺の施設名称を選定する．

その他の実験条件

- 実車に搭載したシステムを使用．
- 車はパーキングなどに停車．エンジンはアイドリング状態とする．
- 発話者はメーカーが推奨する位置で発話を行う．
- エアコン使用．ただし，ファンの動作量は最小設定．
- 窓は全閉．
- オーディオ類は OFF．
- ワイパー，ウインカー等は停止．
- 雨天の実験は避ける．（注：雨音による影響を避けるため．）
- 騒音の大きい場所での評価は避ける．

実験手順

a．現在地設定

現在地によって認識対象を動的に制御する方式をとるカーナビが多いため，カーナビの現在地は“ 新宿，東京都庁（第一本庁舎）前 ”にあらかじめ固定する．

b．音声による POI 入力方法の事前確認

POI をすぐに入力できる機器だけでなく，「施設で探す」などの発話によって，あらかじめ POI 入力を受け付ける状態に遷移しておく必要のある機器もある．事前にマニュアルなどで動作を確認し

ておき，認識性能の評価には，POI 入力を受け付ける状態に設定した後の認識結果のみを使用する．つまり，その状態に遷移するまでに音声入力が必要だったとしても，それまでの性能は評価しない．実験の評価者は，POI 入力を受け付ける状態に移るまでの操作を別途リモコン等で行っておくか，発話者にこの操作を事前に行うことを指導する．

c．発話者の事前訓練

音声入力装置に対する発話者のスキルレベル（習熟度）を統一するため，実験の前に，操作マニュアルを読ませるとともに，各発話者を個々の装置特有の音声入力方法（特に，発話のタイミング）に慣れさせておく．3 箇所程度の POI 入力が正しく行えるまで練習させるのが望ましい．

d．性能評価の手順

先に選定した POI リストから順次一箇所（地名，施設名など）を選び，音声入力を行う．発声の後，入力内容を確認したら，カーナビを音声による POI 入力を再度受け付ける状態に戻し，次の POI の音声入力に移る．これをリスト中の POI すべてに対して繰り返す．全話者の合計発声数に対し，POI の入力が正しく行えたと考えられる数を数え，次式で求まる値をこの装置の認識率とする．

$$\text{認識率}\ (\%) = 100 * \frac{\text{正解総数}}{\text{発話総数}}$$

ただし，正解・不正解は次の基準に従って判断する．

1. 各地名は 1 回のみ発話するものとし，応答が返らない場合や，再発話を促す応答が返る場合も不正解として数える．

2. 発話と表示結果やコールバックが一致しない場合（たとえば，“ TDL ”という発話に対して，“ 東京ディズニーランド ”という正式名称が返される場合など）でも，明らかに目的を達成していると考えられる場合は正解として数える．

3. 発話内容に多義性が残る場合（たとえば，“ 新宿駅 ”という発話に対しては JR 新宿駅，小田急新宿駅，新宿駅東口，新宿駅南口など複数の正解候補がありうるが，その中の 1 箇所が選択されるか，候補リストが返される場合など）も，入力目的を達成していると判断した場合は正解と数える．

4. ”東京タワー展望台“ という発話に対して ”東京タワー蝋人形館“ と認識されれば，これは地図上の場所としては一致するものの，明らかに入力しようとした施設とは異なるので不正解とする．

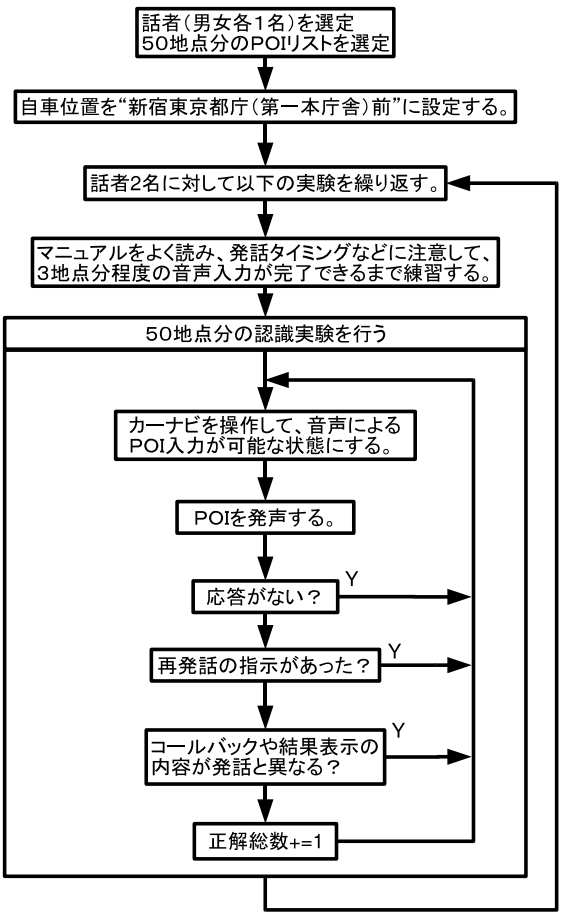

図 1: 簡易評価の手順

5. 使用するカーナビによっては，認識された POI が正しくても，たとえば“ 東京駅 ”という発話が，“ 東京駅を表示する ”，“ 東京駅へ行く ”などのフレーズに誤認識されているケースも考えられうるが，認識内容の詳細を評価者が検証することは困難なので，入力目的を達したものとして，正解に数えてよいものとする．

以上をまとめた手順を図 1 に示す．

3.3.2 推奨評価方法

発話者・被験者

発話者は日本語を母国語とする男女各 10 名（以上）とする．

発話リスト

添付（予定）の POI リストなどを参考に，各話者 100 箇所（以上）を選定する．

その他の実験条件

簡易評価と同じ.

実験手順

a．現在地設定
簡易評価と同じ.
b．音声による POI 入力方法の事前確認
簡易評価と同じ.
c．発話者の事前訓練
簡易評価と同じ.
d．性能評価の手順
簡易評価方法と以下の 2 点を除いて同じである.

1. 使用するカーナビによっては，認識された POI が正しくても，たとえば,“ 東京駅 ”という発話が,“ 東京駅を表示する ”,“ 東京駅へ行く ” などのフレーズに誤認識されているケースも考えられる．**詳細な認識内容が異なることが分かる場合，これらは誤認識として数える．**
2. 可能な限り**実験時の語彙リストサイズを認識率に併記する．**ここでリストサイズとは，評価発声の時点で孤立単語発話で入力可能な項目の総数とする．つまり，階層化された“ 県名＋施設名 ”などの発話は，認識対象であっても語彙には含めない[3].

   なお，発話によってリストサイズが異なるような実験を行う場合は，サイズ s_i の語彙リストに対する発話回数を m_i とすると，以下の加重平均値をこの実験のリストサイズ S とする[4].

$$S = \frac{\sum_i (m_i * s_i)}{\sum_i m_i}$$

3.4 今後の課題

現在，本ガイドラインに添付する POI リストを作成するために，各カーナビメーカーの製品で入力可能な POI や認識対象語の比較検討を行っている.

今後は本ガイドラインを用いた評価を実際に行い，特に以下の項目について検討する必要がある.

- 提案手法が，音声認識の本来持つ能力や利点を引き出せる評価手法となっているか.
- 各メーカの製品の仕様の違いを吸収して，音声入力の性能のみに注目した評価が可能となっているか.
- 特に簡易評価方法において，所要時間や被験者の労力に関して，過大な負担を要さないものとなっているか.
- 特に推奨評価方法において，統計的に意味のある評価が可能であるか．また，絶対値として意味のある値が得られるか.
- ガイドラインの説明は十分にわかりやすいか.

4 まとめ

本報告では，情報処理学会試行標準委員会の WG4 小委員会（音声言語インタフェース）が策定した，(A) ディクテーションに用いる基本記号に対応する読みの試行標準，および (B) カーナビ用音声入力の性能評価ガイドラインの試行標準（案）, の内容を紹介した．本報告は音声言語処理研究・開発に携わる多くの方々の意見を収集することを目的にまとめたものである．今後もメーリングリスト等を通して関連分野の研究者・技術者の方々の意見を反映させ，有用な試行標準の提供を行っていきたい.

参考文献

[1] 新田恒雄, 石川 泰, 伊藤克亘, 畑岡信夫, 松浦博, 磯谷亮輔, 西村雅史, 西本卓也: “音声言語情報処理に関する情報処理学会の試行標準策定活動,” 情処研報 2002-SLP-40-10, pp. 57–60, Feb 2002.

[2] 松浦 博, 西本卓也, 金子 宏, 磯谷亮輔, 石川 泰, 西村雅史, 伊藤克亘, 新田恒雄: “音声認識読み記号および音声関連ソフトウエアに係わる用語の試行標準案,” 情処研報 2003-SLP-45-11, pp. 65–70, Feb 2003.

[3] 情報処理学会情報規格調査会 : “音声認識のための読み表記,” 情報処理学会試行標準 IPSJ-TS 0004:2003, http://www.itscj.ipsj.or.jp/ipsj-ts/02-04/toc.htm

[4] 情報処理学会情報規格調査会 : “ディクテーションに用いる基本記号に対応する読み,” 情報処理学会試行標準 IPSJ-TS 0009:2004, http://www.itscj.ipsj.or.jp/ipsj-ts/ts0009/toc.htm

[3]語彙サイズを並記する目的は，POI 入力タスクの難易度を示すことで，難易度の異なるシステム同士の単純な性能比較を防ぐことである．市町村名などの長い単語列を含めると語彙サイズは極端に大きくなるが，孤立単語を長い単語列に誤認識することは少ないと考えられるため，POI 以外の対象語句を除外するのが妥当である.

[4]見掛けの語彙サイズを増やすために，小語彙タスク 99 回と（誤認識覚悟での）大語彙タスク 1 回を実施する，といった懸念があるため，今後さらなる検討が必要である.