2007－SLP－68　(7)-2
2007／10／19

# 自動車運転行動中発話の分析

旭化成株式会社　音声ソリューションビジネス推進部
庄境 誠，加藤 智之，岡本 淳

**アブストラクト**

カーナビなどの車載情報機器のハンズフリー音声インタフェースの搭載率は高まっているが，利用率が高いとはいえない。利用率に相関の高い実環境性能すら，明らかになっていないのが実情である。そこで，260 名分の自動車運転中発話コーパスを新たに収集し，低認識性能の被験者に注目して行った物理量と認識性能との相関分析結果について論じる。

# Analysis of Utterance during Action of Driving Car

Speech Solutions, Asahi Kasei Corporation
Makoto Shozakai, Tomoyuki Kato and Jun Okamoto

**Abstract**

Although a loading rate of hands-free voice user interface capability into in-vehicle information appliances such as GPS navigation system has become high, a usage rate of the capability is not necessarily high. Even performance in adverse environments, which is highly correlated to the usage rate, is not revealed yet. We collected the speech corpus of utterance of 260 persons during action of driving car. Results of correlation analysis between acoustic parameters and performances for low performance subjects are discussed.

## 1. 背 景

音声認識のアプリケーションの中で，ハンズフリーインタフェースの必要性が明確な車載情報機器（カーナビ，カーオーディオ，携帯電話）への音声インタフェースの搭載率は，他の潜在的アプリケーションに対して高いと言える。しかしながら，それらの機器のハンズフリー音声インタフェースの使用率は必ずしも高いとは言えないのが実情である。その原因についても，様々な議論がなされてきた。

万人に使用される車載情報機器のインタフェースとして期待される役割を考えれば、様々な未解決課題の中でも，実環境における十分高い対雑音性能，対話者性能，対語彙性能の確保の問題を無視するわけにはいかない。本論では，この問題について，議論する。

## 2. 認識性能の表現

実環境における対雑音性能，対話者性能，対語彙性能には分布（ばらつき）が存在する。例えば，認識性能に関して降順に話者（利用者，被験者）を横軸に並べた場合，縦軸の認識性能は図 1 のような形状を取る場合が一般的である。ここで，A 点(*, y)は，全話者の平均性能を表す。B 点(0, y)は，全話者の最高認識性能を表す。W 点(100, y)は全話者の最低認識性能を表す。I 点(x, y)は，認識性能分布の屈曲(Inflexion)点を表す。認識タスクにも依るが，B 点から I 点まではなだらかに認識性能が低下し，I 点から W 点までは急激に認識性能が低下する傾向が一般的である。

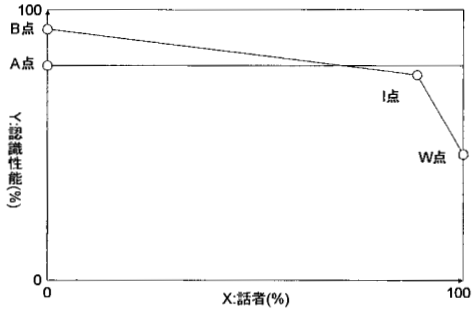


図 1 認識性能分布の概形

従来，認識性能は平均認識性能である A 点の y 値のみで表現されることが多かった。大語彙連続音声認識のような難しい認識タスクの場合も，A 点の 100-y 値（誤り率）がどのくらい改善したという研究が多い。その一方で，W 点や I 点の周辺に着目し，それらの y 値が 100 から乖離している原因の本質に迫る研究の取り組みがあまり見られないのが実情である。

一方，音声認識機能を実用化の場面で使う側の立場で，音声認識技術を外の世界から眺めた場合に，まず関心があるのは，W

点のｙ値である。この値は，実環境での認識性能の品質保証（最低保証）に直結するからである。しかし，実環境においては，Ｗ点のｙ値が90(%)を大きく下回る場合も多い。その場合でも，Ｉ点の(x,y)値に着目し，それらが(80,90)，(90,90)，(95,95)などの値以上であることが確認できれば，製品・サービスの企画コンセプト次第では実用化に踏み切るという判断もあろう。

3．自動車運転行動中発話コーパスの収集

自動車運転行動中の発話における対雑音性能，対話者性能，対語彙性能を明らかにするためには，それらの分析が可能な発話コーパスの確保が必要である。自動車運転行動中の発話コーパスとしては，CIAIRコーパス[1]が知られているが，被験者の発話内容が完全に制御されておらず，対雑音性能，対話者性能，対語彙性能を客観的に分析するには好適ではなかった。そこで，以下の条件を満足する自動車運転行動中の発話コーパス(260名)を新たに収集した。

1）録音ゲインが全収録で固定

2）発話内容が複数の語彙種を含んでおり，全話者で共通

3）発話の際の自動車運転行動が，全話者で共通

これにより，雑音性，話者性，語彙性を切り分けて分析できる。収集コーパスの概要を表1に，収録条件を表2に示す。

表1 収集コーパスの概要

| 被験者種別 | 走行状態 | 人数 | 目的 |
|---|---|---|---|
| 一般被験者（一般人） | アイドリング/コース内走行 | 男女各110名 | 日本語発声の話者による広がりを把握 |
| プロ被験者（運転のプロ） | アイドリング/市街地走行/高速走行 | 男女各20名 | より現実的な運転行動中音声を把握 |

表2 収録条件

| 標本化周波数 | ビット長 | チャネル数 | 被験者への語彙の指示 | 使用車両 |
|---|---|---|---|---|
| 48kHz | 24bit | 4ch | TTS | Vitz(TOYOTA) |

4．発話コーパスの物理量分析

収集した発話コーパスの対話者性能を解析するために，以下の物理量の統計分析を行った[2]。

(1)音圧：SS後の10msフレーム最大平均パワー

(2)SNNR：プリエンファシス後の発話/雑音区間のパワー比(dB)

(3)発声速度：単位時間当たりのモーラ数(値が大きいと早口)

(4)滑舌度：特定話者音響モデルの5母音間距離(値が小さいと滑舌が悪いといえる)

(5)話者固有度：不特定話者音響モデルに対する特定話者音響モデルの距離(値が大きいとスペクトルの固有度が大きく，特徴的な声といえる)

5．低認識率被験者の物理量と認識性能の相関分析

表3の認識タスク条件で求めた認識性能の分布を図2に示す。図2のＷ点，Ｉ点付近の丸で囲まれた2名の女性プロ被験者の物理量のヒストグラムを全女性プロ被験者の物理量のヒストグラムの上に重畳表示した結果を図3に示す。2名の被験者の特徴的な物理量を楕円で囲んで示す。SNNR，滑舌度，話者固有度が低認識性能に強い相関を持つことが示唆される。

表3 認識タスク条件

| | | |
|---|---|---|
| 音響モデル | Monophone/43音素/単一正規分布 | Leave-one-outによる女性不特定話者モデル |
| 待ち受け語彙 | 約1000単語 | カーナビ用想定コマンド |
| 使用音声 | 女性プロ被験者(20名)/高速走行中音声/マップランプ付近設置マイク収録音声(1ch) | |
| 特徴量 | MFCC(12)，ΔMFCC(12)，ΔlogPower(1) | サンプリング周波数16kHz |

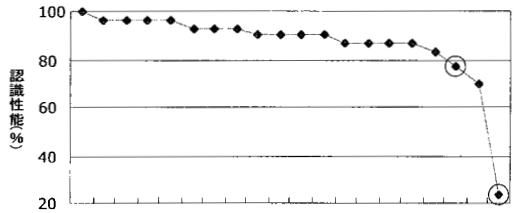


図2 女性プロ被験者の認識性能分布

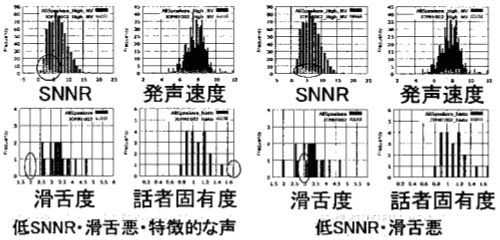


図3 Ｗ点(左)，Ｉ点付近(右)の被験者の物理量分析

6．今後の予定

低性能や誤認識の原因は多様である。上述の相関分析を詳細に進め，対雑音性能，対話者性能，対語彙性能の低下に対して，より大きな影響を与える原因を特定し，順次，対策を講じることにより，Ｉ点の(x,y)値が(100,100)に近づく効果を定量化して行く予定である。

参考文献

[1] 河口，松原，岩，梶田，武田，板倉：実走行車内における音声データベースの構築，情報処理学会，音声言語情報処理研究会，SLP-30-12，2000.

[2] 加藤，岡本，庄境：自動車運転行動中発話の日本語音声コーパスの物理量と認識性能の相関分析，日本音響学会，音講論，3-Q-25，2007.9.