

4. 立体音メニューを用いた音声呈示の検討

4.1 本章の目的

音声対話に関する基礎技術研究では従来の音声処理に加えて自然言語処理・知識処理・対話モデルなどの種々の技術を統合した方式が検討されている [4-1]. 自然言語処理の分野では, 人間の概念をどのように機械に理解させるかを, 言語理解処理を用いて実現しようとしている例が挙げられる [4-2]. 知識処理の分野では, 文中の抽象名詞が他の抽象名詞と交替する現象に注目し, これをニューラルネットワークの一手法である自己組織化マップを用いて解析する手法が提案されている [4-3]. 対話モデルの分野では, 人間同士の対話が成立するための前提をモデル化し, 省略や倒置表現を対話管理に有効な情報として活用する試みが行われている [4-4]. しかし, これらはいずれも基礎研究の途上であり, 実用段階にはまだ達していない. したがって, 市販製品レベルでは情報機器から音声ガイダンスがあり, それに呼応する形でマイクに向かって短い音声コマンドをドライバが発話する, 逐次音声操作方式が主流である.

本研究が対象とする応用分野において音声認識技術を実用化する上で, 現実問題としてUIをもっと使いやすいものにすべきという指摘がある [4-5]. 使いやすさに影響を与えている問題の一つとしては, 音声誤認識が発生した状況に対するユーザの認識不足が考えられる. 例えば第3章のユーザテストでは, 平均5.4音節の長さをもった音声コマンドによる逐次音声操作を行っている際, 操作系が切り替わる箇所で音声誤認識が発生し, その復唱内容をユーザが聞き間違えたか聞き逃して操作を進めた結果, タスクが失敗に終わってしまう問題が報告されている.

この問題の解決法として, 「xでよろしいですね?」と復唱して確認をとる方法が考えられる. しかし, この方法では, 音声認識結果が1種類のみ呈示されるので, 音声誤認識が起きたときには手順をいちいち戻ってやりなおさなければならず, ユーザに心理的な抵抗を与える可能性がある. そこで, 本研究では, 複数の音声認識候補を音声呈示してユーザに選択させる 1) メニュー方式, および音声情報の内容を空間的な位置情報として意味づける 2) 立体音方式の2種類のユーザインタフェースを提案する. そして, これらの方式が運転行動および車載情報機器の逐次音声操作を同時に行うデュアルタスク環境において運転パフォーマンスおよび音声誤認識発生時における復唱音声の理解に与える影響について検討を行った.

4.2 デュアルタスク環境

4.2.1 デュアルタスク環境における運転余裕度モデル

本研究が対象とするデュアルタスク環境における認知モデルについて概観する。人間の情報処理リソース（以下、資源）の容量を一定とすると、図 4.1 に示すように、あるタスクを行うことにより負荷に応じた処理資源が消費されるため、残された余裕容量が減ることになる [4-6]。ここで二種類のタスクを同時に実施するデュアルタスク環境を想定すると、一方のタスクにより余裕容量が減るため、他方のタスクに割り当てることができる処理資源が減少し、結果として遂行成績は低下すると予想される。運転中の注意分散とは、ドライバが運転とは直接関係ない事物へ注意を向けたために余裕容量が減じた状態であり、残された余裕容量の大きさに応じて運転行動に影響が生じる。

このモデル概念に基づき、本研究ではデュアルタスク法を用いて検討を進めた。デュアルタスク法とは、二重課題法とも呼ばれ、2 つのタスクを同時に行ったときに、主たるタスク（主タスク）に対してもう一方の副次的なタスク（副次タスク）の遂行が与える影響を計測する手法である。本研究では自動車運転環境を対象とするため、主タスクは運転行動、副次タスクは車載情報機器操作と位置づけた。そして、副次タスクにおける呈示方式の違いによる認知的負荷の差を運転中の注意分散ととらえ、これが主タスクである運転行動に与える影響を見ることができる。さらに、副次タスク遂行に必要な資源の消費状況を相対的に見ることができると考えられる。

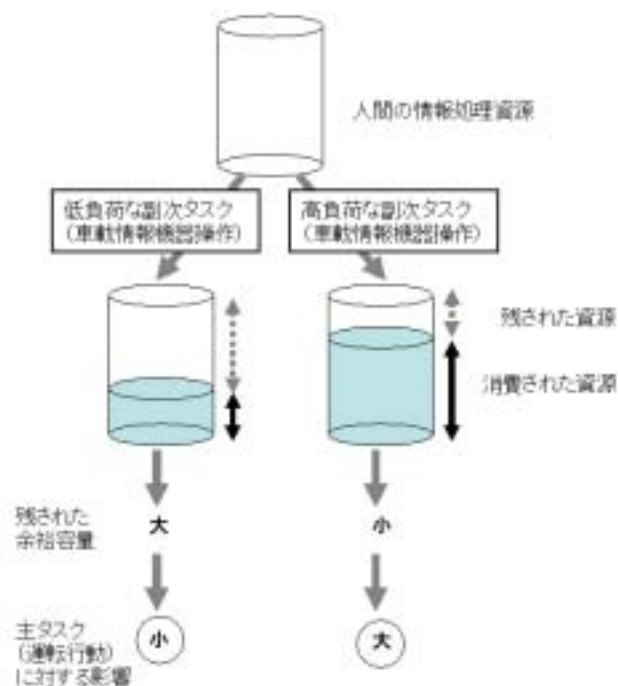


図 4.1 デュアルタスクモデルにおけるタスク請求と運転余裕容量の変化

4.3 主タスク

4.3.1 運転シミュレータの概要

主タスク用の運転シミュレータは C++言語を用いて開発し PC (OS: Windows 2000) 上で動作する。市街地道路を想定し, CG 映像を用いた直線道路, 建物, 歩道, 街路樹, 前方車両ならびに対向車両から構成される (図 4.2)。映像は 100 インチスクリーンに投影して呈示した。このシミュレータには車速制御の機能をもたせた。具体的にはキーボードの「」キーを押すことで車速を制御する。車速は 3 段階のスピード (30km/h, 48km/h, 72km/h) で, 「」キーを 1 秒間押すと 1 段階高速に, 1 秒間離すと 1 段階低速となる。速度変化は 15Hz で標本化され, セッション終了時に自動的にログファイルとして書き出される。ロードノイズは普通乗用車の車内において録音したものをスピードに合わせて 3 段階に変化させるようプログラムした。被験者にはヘッドホンを通じて呈示した。音圧は 60dB (A) とした。



図 4.2 シミュレータ画面

4.3.2 主タスクの設定

主タスクは運転シミュレータ上での擬似運転行動であった。本研究では擬似運転行動は速度制御のみとした。この他にハンドル操作が考えられるが, 道路形状による影響を排除するために今回は速度のみを考慮することとした。速度の制御は認知負荷をできるだけ一定にするため, 運転中は 48km/h を保つ努力を求めた。この速度を保つためには, 約 1 秒おきに「」キーの押し・離しを頻繁に繰り返さなければならない。

4.4 副次タスク

4.4.1 音声操作シミュレータの概要

副次タスクに使用する音声操作シミュレータの構築を行った。本シミュレータは Director8.5 (Macromedia)を用いて開発し, PC (OS: Windows 2000) 上で動作する。音声を用いた逐次対話は, 図 4.3 に示す形で定義した。まず, システムよりガイダンスがあり,

それに応じて、ユーザは発話許可ボタンを押した後、ピップ音が聞こえたら短い音声コマンドを発話する。そして入力されたコマンドは音声認識され、その結果を元に次のステップのガイダンスを呈示する、というながれとなっている。また、音声誤認識が発生した際には、前のステップに戻れるよう、復帰ボタンを用意した。発話許可ボタンは音声操作シミュレータ用キーボード上の「」キーに、復帰ボタンは「」キーにそれぞれ割り当てた。本シミュレータには目的地設定の機能をもたせた。設定は音声ガイダンスに従って設定モード名、目的地の設定方法、施設の種類名、施設のある都道府県名、そして施設の固有名称の順番に入力するながれとした。

音声認識精度を厳密に統制するため、実際には音声認識エンジンの実装は行わず、音声の入力を感知すると、あらかじめプログラムされた応答音声呈示される。これにより、被験者には音声認識が行われているかのように見える。実験タスク統制の関係から、図 4.4 に示すように、音声誤認識は決まった箇所発生するように設定された。音声誤認識が発生したことをユーザが気づけば、その時点で復帰ボタンを押すことで作業のやりなおしができる。しかし、音声誤認識を合計 2 回見逃してしまうと、自動的に行き先設定が失敗に終わるアルゴリズムとした。これにより、音声誤認識の発生を正しく理解したことを達成率という形で測定することができる。音声は合成音声 (Pro Talker/BM) の女声を使用し、ヘッドホンを通じて 0.4 秒/音節の速さで被験者に呈示した。音圧は 65dB (A) とした。

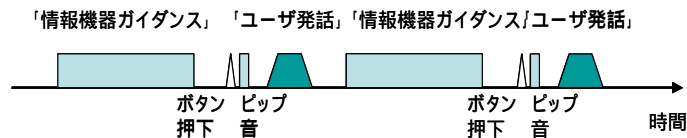


図 4.3 逐次音声操作型システムにおける人間とシステムの対話

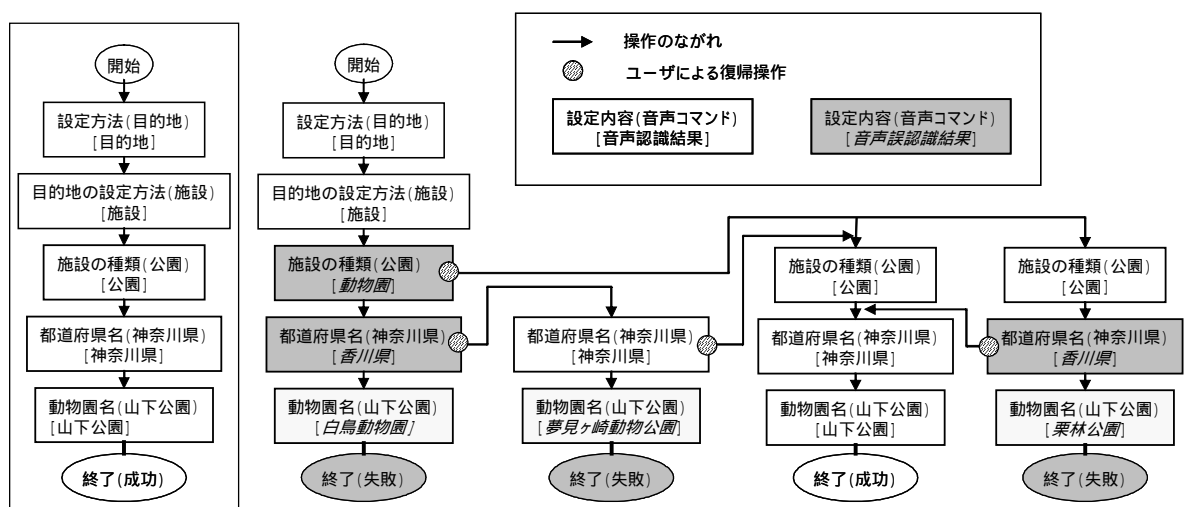


図 4.4 音声操作シミュレータの操作フローとタスクの成否設定

4.4.2 副次タスクの設定

副次タスクは、車載情報機器を音声操作して目的地設定を行うことであった。音声操作シミュレータからの問いかけに対して、短い音声コマンドを入力することで設定を進めていく。このとき発話するコマンドは、設定モード名に「目的地」、目的地の設定に「施設」、施設の種類名に「公園」、施設のある都道府県名に「神奈川県」、そして施設の固有名称に「山下公園」を用いた。

4.5 実験条件群

副次タスクで呈示する方式として以下の5群を設定した。各群における人間とシステム間の対話を時系列的に示したものが図4.5である。

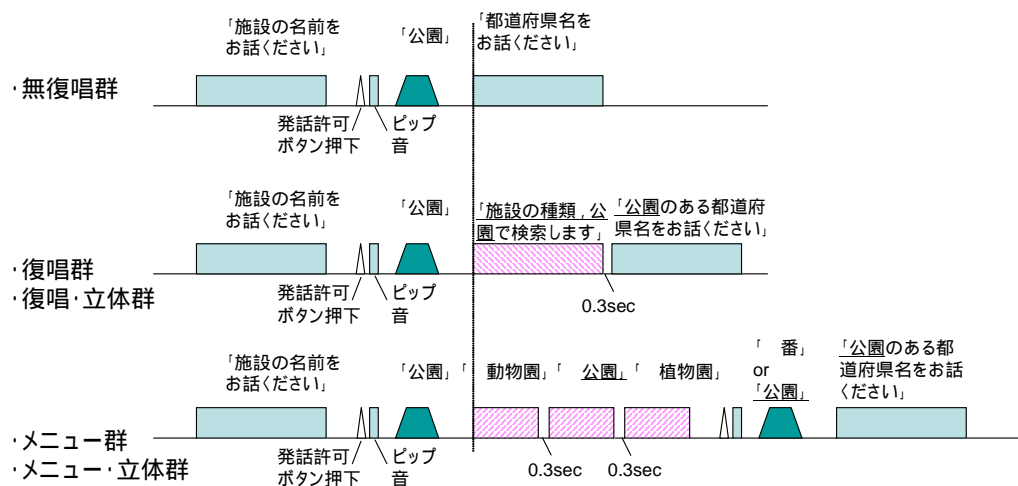


図 4.5 音声操作シミュレータにおける各群の呈示方式

(1) 無復唱群

無復唱群は、音声認識結果を情報機器が音声復唱することなく、次のガイダンスに進む条件である。これは一部の市販製品にみられる方式である。音声は合成音声で、ヘッドホンを介してモノラル呈示される。

(2) 復唱群

復唱群は、音声認識結果を音声復唱という形でフィードバックする条件である。この条件では、次のガイダンスに進む前に、現在の操作系と音声認識結果の両方を音声呈示する。音声は合成音声で、ヘッドホンを介してモノラル呈示される。

(3) メニュー群

メニュー群は、音声認識結果の候補単語を3つ準備し、メニューとして0.3秒間隔で順

次呈示する．その中にユーザの意図するコマンドがある場合はそれを選択し，なければ 1 階層戻ってやりなおす条件とした．音声は合成音声で，ヘッドホンを通してモノラル呈示される．メニュー型の UI は，タスクやコマンド選択のために GUI で最も多用されている方法である [4-7]．この手法は複数のタスクやコマンドで関係のあるものを並べておき，そのなかから任意の情報を選択する仕組みになっている．これを認識候補の呈示に用いることで，複数の選択肢から正しい結果をユーザが意図的に選択でき，見た目の音声認識性能向上を図ることができると考えた．また，音声コマンドを発話することは現在階層を理解するうえで有用であるという報告 [4-8] がある．このことから，ユーザの注意が選択肢に向けられることで，それが操作系に関わる単語だった場合，現在操作系の把握に貢献すると考えた．

(4) 復唱・立体群

復唱・立体群では，復唱群で用いた復唱音声を左前方 60° 距離 80cm に音像として配置した．その後流れるガイダンス音声は右前方 60° 距離 80cm の位置に音像を配置した．位置情報以外の音声呈示内容，タイミングなどは復唱群と同条件とした．音像の立体的な位置情報は 3 次元音場を収録可能なダミーヘッドマイクと呼ばれる人型の収録装置に合成音声をスピーカ呈示し，PC にバイノーラル収録することで得た．このとき量子化 bit 数 16bit，標本化周波数 44.1kHz とした．呈示はヘッドホンを用いた．本条件の設定は，呈示音声に立体的な位置情報を付加することで記憶成績が向上するという報告 [4-9] に基づいている．具体的には，シミュレータ運転環境において模擬運転行動中に，複数のカテゴリの単語を系列的に音声呈示してできるだけ多く記憶させ，呈示終了後に自由再生させる課題を課したところ，単語のカテゴリごとに音源位置を統一して与えた条件では，ランダムな位置から与えた条件に比べて記憶成績が顕著に向上した結果に基づいている．

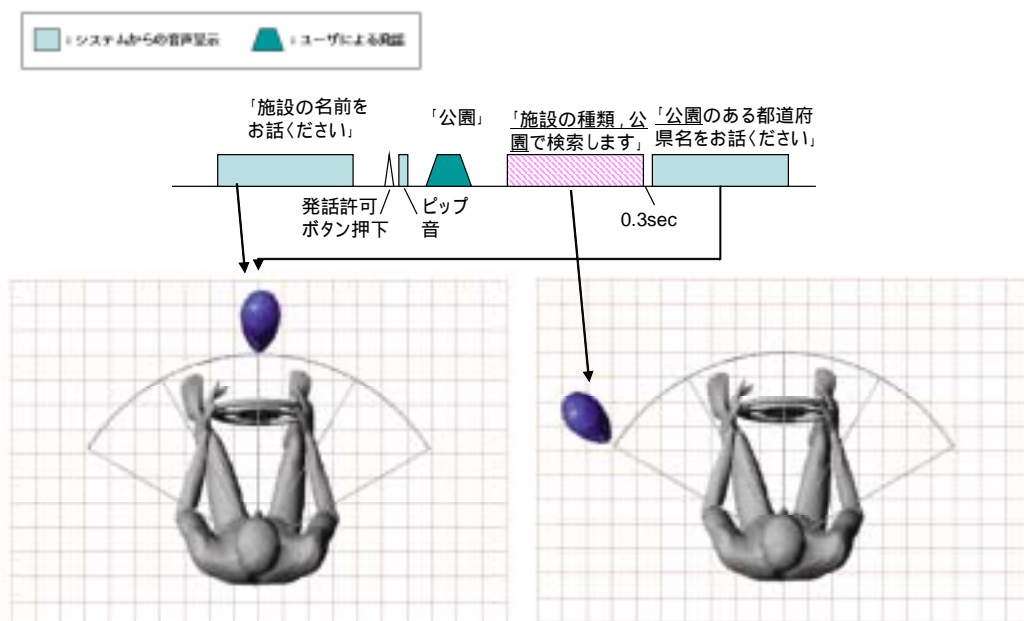


図 4.6 復唱・立体群呈示のながれと音源配置

(5) メニュー・立体群

メニュー条件と、立体的な位置情報の2条件を組み合わせた群を設定した。音声認識結果の3候補は前方左60°、正面、そして前方右60°の位置において、呈示間隔を0.3秒として1候補ずつ順番に呈示した(図5)。メニューの選択肢を3候補としたのは、前述の先行研究[4-9]において音源数が1,3,5と設定した場合、3の場合において最も記憶成績が高かったことを根拠とした。

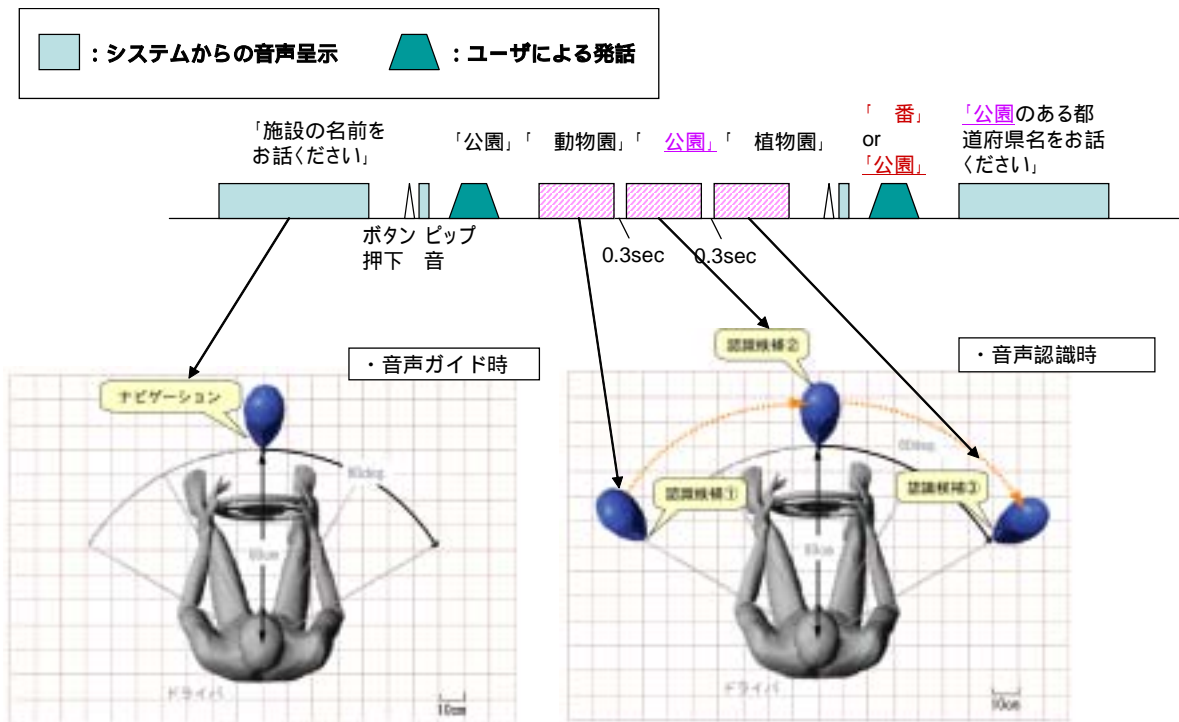


図 4.7 メニュー・立体群呈示のながれと音源配置

立体音響の収録機材は以下のとおりであった。収録の様子を図 4.8 に示す。

- ・ ダミーヘッドマイク (HATS Type 4128 / B&K)
- ・ ダミーヘッドマイク用プリアンプ (Nexus / B&K)
- ・ ハードディスクレコーディング用コンピュータ
(Power Macintosh G4 400MHz / Apple Computer)
- ・ 入出力用サウンドデバイス (Digi 001 / Digidesign)
- ・ 音声出力用スピーカ (SMS-1P / Sony)
- ・ 音声送出力用 PC (Think Pad X22 / IBM)

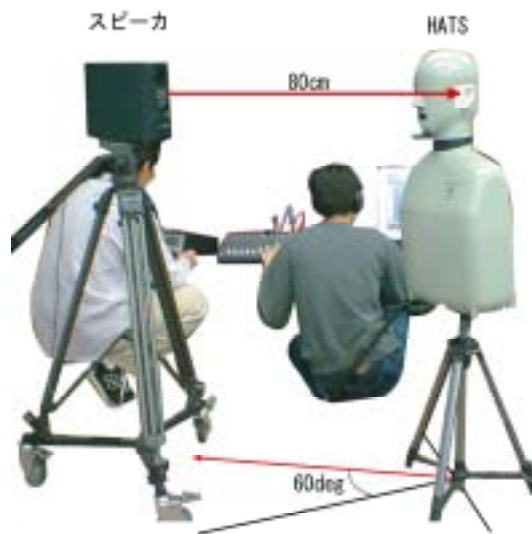


図 4.8 バイノーラル音響収録風景

4.6 被験者

被験者は視聴覚が健常で自動車運転免許をもつ大学生計 38 人（男性 26 人，女性 12 人）であった。年齢は 18 歳から 25 歳，平均 21.4 歳，標準偏差 ± 1.86 歳であった。実験は被験者間計画とした。各群の被験者数はそれぞれ，無復唱群 7 人（男性 5 人，女性 2 人），復唱群 7 人（男性 5 人，女性 2 人），メニュー群 7 人（男性 5 人，女性 2 人），復唱・立体群 8 人（男性 5 人，女性 3 人），メニュー・立体群 9 人（男性 6 人，女性 3 人）であった。

4.7 手続き

実験手続きは以下のとおり行った。なお，実験は被験者 1 人あたり 1 条件で行われた。

- 1) まず，車速を 48km/h に保つための練習走行として，運転プログラムの操作のみ約 5 分間行わせた。
- 2) 練習走行終了後，カーナビの操作法を説明した。このとき，音声コマンドを正しく再生できるまで被験者に復唱させた。
- 3) タスクの遂行にあたり，運転プログラム用のキーボードは右手，カーナビ用は左手で操作するよう教示した（図 4.9）。そしてできるだけ車速を 48km/h に保ち，かつ正確にカーナビの操作をするよう教示した。教示後，デュアルタスクの試行を開始した。試行は 1 人あたり 1 回とした。
- 4) 音声操作シミュレータ操作が終了した時点で実験プログラムを停止し，試行終了とした。
- 5) その直後にインタビューを行い，内省報告を採取した。

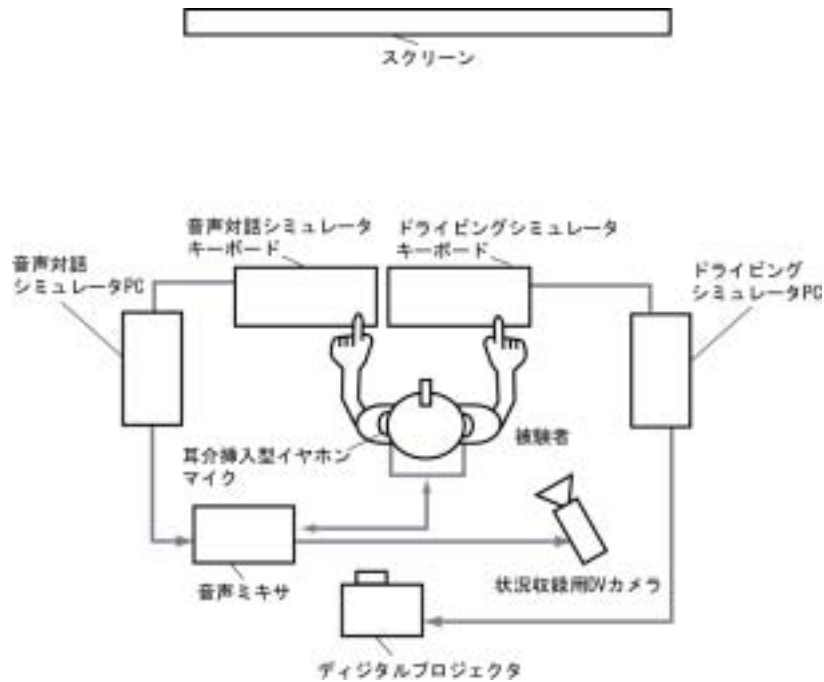


図 4.9 実験システム

4.8 評価手法

評価は(1)目的地設定タスク達成率,(2)運転速度不安定度,および(3)内省報告で行った。

4.8.1 目的地設定タスク達成率

4.4において設定した条件下では,どのような選択を行っても音声誤認識を2度聞き逃すと,自動的にタスクが失敗するよう統制されている。したがって,認識結果およびそれに基づいて呈示されたフィードバック音声をどれだけ正しく理解できたかを達成率で表すことができると考えられる。そこで,目的地設定を最終的に正しく行った場合を達成(100%)とみなして100分率で評価を行った。

4.8.2 運転速度不安定度

デュアルタスク環境において,副次タスクの認知的負荷の差が注意分散という形で運転行動に影響を与えることは既に述べた。運転パフォーマンスに関する先行研究[4-10]では,0.2秒あたりの車間距離の平均変動量と車体の左右方向への平均変動量を標準化したものを運転不安定度とする指標が提案されている。このほかに,速度の変化について最小,平均,変動と歪みの4変数に着目した例もある[4-11]。これらに共通して用いられている速度の変化情報である。したがって,これをひとつの指標として用いることは有用であると考えられる。そこで,本研究では速度制御情報を利用することにした。ただし,速度情報

のみを利用するため、ここでは運転速度不安定度と呼ぶことにする。自動車運転タスクとして、3段階の車速のうち48km/hをできるだけ保つように教示することで、副次タスクの認知的負荷の違いが、車速制御の不安定性に影響すると考えた。そこで、練習走行およびデュアルタスク走行時間における48km/h保持時間の割合を求め、その差を式(4.1)に示す時間運転速度不安定度と定義した。

$$= \frac{\text{練習走行時の48km/h保持時間の合計}}{\text{練習走行時間の合計}} - \frac{\text{デュアルタスク走行時の48km/h保持時間の合計}}{\text{デュアルタスク走行時間の合計}} \quad (4.1)$$

運転速度不安定度値が高いほど、相対的に副次タスク遂行に必要な資源が多いことを表していると考えられる。

4.8.3 内省報告

復唱やガイダンスがどの方向から聞こえたか、どのような意味付けがなされていると考えて聴取したかについて、特に音源の立体的な位置情報を与えた2群を対象にインタビュー形式で内省報告を採取した。これにより、立体的な位置情報を与えたことが主観的にどのような影響を及ぼしているかについて調べた。

4.9 結果

4.9.1 目的地設定タスク達成率

副次タスク達成率の各群における平均を図4.10に示す。無復唱群と復唱群との間では特に差は見られなかった。一方、メニュー群の成績は復唱群と比して8.3ポイントと若干高い結果となった。また、復唱・立体群、メニュー・立体群いずれの場合も100%の達成率であった。

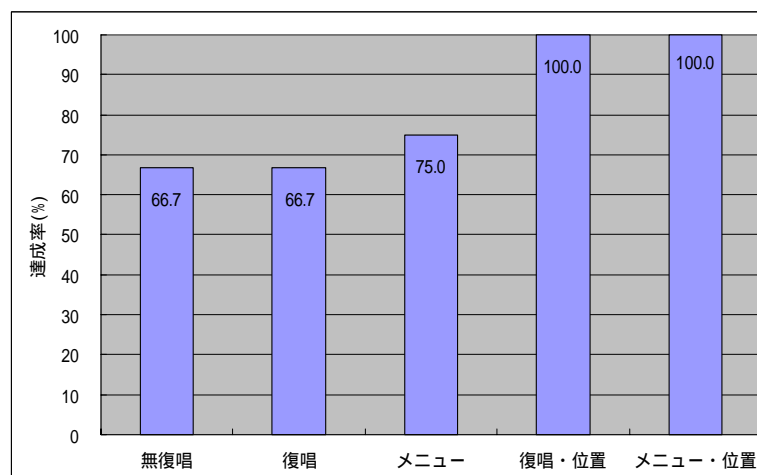


図 4.10 目的地設定タスク達成率

4.9.2 運転速度不安定度

各被験者の運転速度不安定度値を求め、これに対して一元配置の分散分析を行った。その結果、条件の効果が有意であった ($F(4, 33) = 3.43, p < .05$)。そこで、LSD 法による多重比較を行った。その結果、復唱無群 < メニュー群 ($MSe = 0.0086, 5\%$ 水準), 復唱群 < メニュー群 ($MSe = 0.0086, 5\%$ 水準), メニュー群 > メニュー・立体群 ($MSe = 0.0086, 5\%$ 水準) という関係が明らかとなった (図 4.11)。しかし、メニュー群と復唱・立体群の間には有意差は認められなかった。図 4.11 は各群における結果を箱ヒゲ図で示したもので、箱の下辺が 25 パーセンタイル、真ん中の線が 50 パーセンタイル (中央値), 上辺が 75 パーセンタイルを示している。

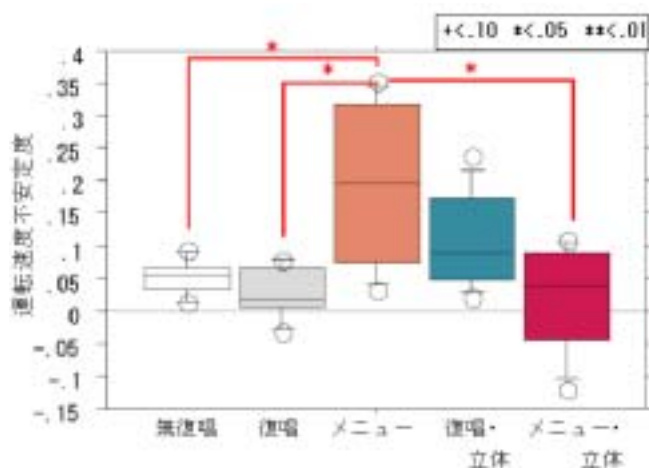


図 4.11 運転速度不安定度

4.9.3 内省報告

(1) 復唱・位置群

内省報告より、復唱とガイダンス音声それぞれ左右から聞こえてくると答えられた被験者は 18.2% だった。さらに複数の方向から聞こえてきたが意味がはっきりとわからず、混乱したことを示すプロトコルが 72.7% の被験者から採取された。

(2) メニュー・位置群

内省報告より、メニューが 3 方向から聞こえてくること、ガイダンスが中央から聞こえてくることを正しく答えられた被験者は群内において 87.5% であった。また、特に音源の位置情報によって混乱をきたしたことを示唆するプロトコルは得られなかった。

4.10 考察

先行研究 [4-12] によれば他作業を行いながらメッセージを聴取し、かつ再生までに時

間が経過する場合に、1回の聴取にて正しく内容を再生できるのは12音節以下のメッセージといわれている。さらに、同研究においては持続時間が4.0秒(24音節)を上回るメッセージについては、1回の聴取では十分に情報の取得と保持ができない場合があると指摘されている。今回の音声呈示における話速は先行研究(0.17秒/音節)と大差ない0.14秒/音節であったので比較が可能である。その場合、メニュー群の不安定度が他の群より顕著に高かった理由としては、復唱部分とガイダンス部分を合わせて平均21.3音節と24音節に近い数値になっていたことが考えられる。これに加え、音声メニューという、慣れない情報呈示が行われたために、聴取と保持に情報処理の資源を大きく配分する必要があったのではないかと考えられる。

一方、同じ音節数条件にもかかわらず、メニュー・立体群は達成率、運転速度不安定度ともに、顕著に改善されている。これは、内省報告結果より、設計時に意図した情報整理が問題なく受け入れられた結果、認知処理的な情報の「精緻化」や「体制化」[4-13]が行われ、処理水準を深めることができたために、情報の聴取と記憶の保持が効果的に働いたのではないかと考察される。このことは、別の呈示法である復唱方式の場合も、位置情報を加えることが達成率向上につながったと思われる。ただし、その説明だけでは復唱・立体群における運転速度不安定度が比較的高くなったことを説明できない。この点については、内省報告で得られた、80%以上の被験者が音源の位置情報と呈示される音声情報との関係付けに失敗している点で説明できる可能性がある。つまり、ベースとなる呈示方式に抛らず音源に位置情報を与えることは、認知的な情報処理特性の向上に寄与する可能性がある。しかし、音源の位置情報と呈示される音声情報の内容との関連付けをユーザが理解できることが重要になってくると考えられる。

4.11 まとめ

本研究では、複数の音声認識候補を音声呈示してユーザに選択させる1)メニュー方式、および音声情報の内容を空間的な位置情報として意味づける2)立体音方式の2種類のユーザインタフェースを提案した。そして、これらの方式が運転行動および車載情報機器の逐次音声操作を同時に行うデュアルタスク環境において、運転パフォーマンスおよび音声誤認識発生時における復唱音声の理解に与える影響について検討を行った。その結果、以下の知見が得られた。

(1) 今回の条件では音声認識結果の復唱は行わない場合に比べて運転への影響が若干減少する程度だが、音源の立体的な位置情報を加えると、復唱音声への理解が進み、状況把握の正確性が向上する。

(2) ただし、ユーザが音源の位置と音源を分けた意味を対応づけられない場合、運転に必要な人間の情報処理資源を情報機器操作にまわしてしまうことがあり、運転に対する負

の影響がみられる。

(3) 今回の条件において音声認識候補のメニュー呈示は、一次元的に呈示すると運転に対して顕著な負の影響を与えるので好ましくない。

(4) ただし、音源の立体的な位置情報を加えることによってユーザにおける情報の整理が促進され、音声誤認識発生状況の把握が容易になる。その結果、運転に対しての低影響化が図られる。