非公開

TR-I-0142

種々の音韻連接単位を用いる規則合成方式の診断的な評価

Evaluation and Diagnosis of Selective Use of Non-uniform Units for Speech Synthesis-by-rule.

武田 一哉, 安部勝雄, 匂坂 芳典
Kazuya TAKEDA, Katsuo ABE and Yoshinori SAGISAKA

1990.2

内容更概

種々の合成単位を選択的に用いる規則合成システムの評価実験を行った。さらに、実験結果に基づいて、種々の単位使用条件と、その結果得られた合成音声の品質との関係を分析した。これらの結果から、合成単位の選択的な使用が優れた合成音を生成することを明らかにするとともに、合成単位の使用条件と得られる品質との関係を幾つかの視点から明らかにした。

ATR 自動翻訳電話研究所
ATR Interpreting Telephony Research Laboratories

1 はじめに

単位接続型の規則合成においては、単位の使用法(どのような単位を、どのように抽出して用いるか)が、大きく合成音声の品質を左右する。これまで、多くの基本音声単位が提案され、CV、VCV,CVCといった現在多くの実用システムが用いている、音韻複合単位が開発されるに至った[1],[2],[3]。一方、これらの基本単位に対応する音声単位素片の作成法については、システムの性能向上に欠くことの出来ない基本要素であるにもかかわらず、未だ系統的な研究がなされていない。この問題は、システムに固有の問題としてとらえられ、種々の基本単位に適用可能な一般的な単位抽出法を確立する努力はなされていない[4],[5]。さらに、いくつかのバリエーションを持つ音声単位を用意し、コンテクストに応じてこれらを選択的に使用するといった、単位の選択的な使用の導入については、ようやく研究がはじめられた段階である[6],[7]。即ち、より自由度の高い単位使用を可能とすることで、合成音声の品質を向上させるための研究が、現在最も必要とされている。

そこで本稿では、自由度の高い単位使用に欠くことの出来ない知見である、合成単位の性質と合成音声の品質との関係、を考察する。即ち、種々の合成単位を選択的に用いる合成方式[8]の枠組みの上に実現した規則合成システム[9]、の品質評価の結果から、音節了解誤りの原因を考察し、これらがどのような単位使用の結果生じたかを明らかにする。

以下第2節で合成方式の概要を述べ、第3節でこの方式の基本的な性能を評価する。さらに4節では、音節了解誤りに着目して、合成単位の性質と合成音声の品質との関係を明らかにする。

2 種々の音韻連接単位を用いる合成方式[8]

図1に、実験に用いた合成システムの構成を示す。システムは、以下に述べる音声データベース,音声素片辞書,単位選択部及びLMAフィルタ[10]を用いた合成系から構成されている。

音声データベース[11]

重要語5240語の発声からなる単語データベースであり、分析パラメータとして30次までのケプストラム係数が格納されている[12]。さらに発声には視察に基づく音韻ラベルが付与されている。

音声素片辞書(SUE辞書)

音声データベース中に存在する全ての音韻連鎖のうち、単位として使用可能な連鎖が合成素片候補として予めSUE辞書上に展開されている。即ち、特定の分割により得られる音声素片のみを用いることにより、データベースからの抽出を容易にする方向で、合成単位候補数の低減を図っている。さらにSUE辞書には、当該音声素片の音韻環境が格納されており、当該音声素片を合成単位として使用する際の、切り出しの難易度が知られる。

単位選択部[13]

単位選択部として、2つのアルゴリズムが提案されている。第一の方法は、可能な単位の組合せの中から、1)接続特性、2)単位の抽出/使用環境の類似性、の2点を最も良く満たす系列を、動的計画法により選択する方法である(DP法)。第二の方法は、入力音韻系列を接続環境に着目して、部分音韻列に分割し、得られた単位候補の中から、最適な単位系列を決定する方法である(EXP法)。

合成部

選択された音声素片区間の音声パラメータ(30次ケプストラム係数)が、パラメータファイルから音韻ラベルに基づいて抽出され、合成単位として用いられる。単位間の接続に際しては、音韻環境に応じてケプストラム距離に基づいた最適な接続点を探索することにより、歪の低減を図っている[14]。

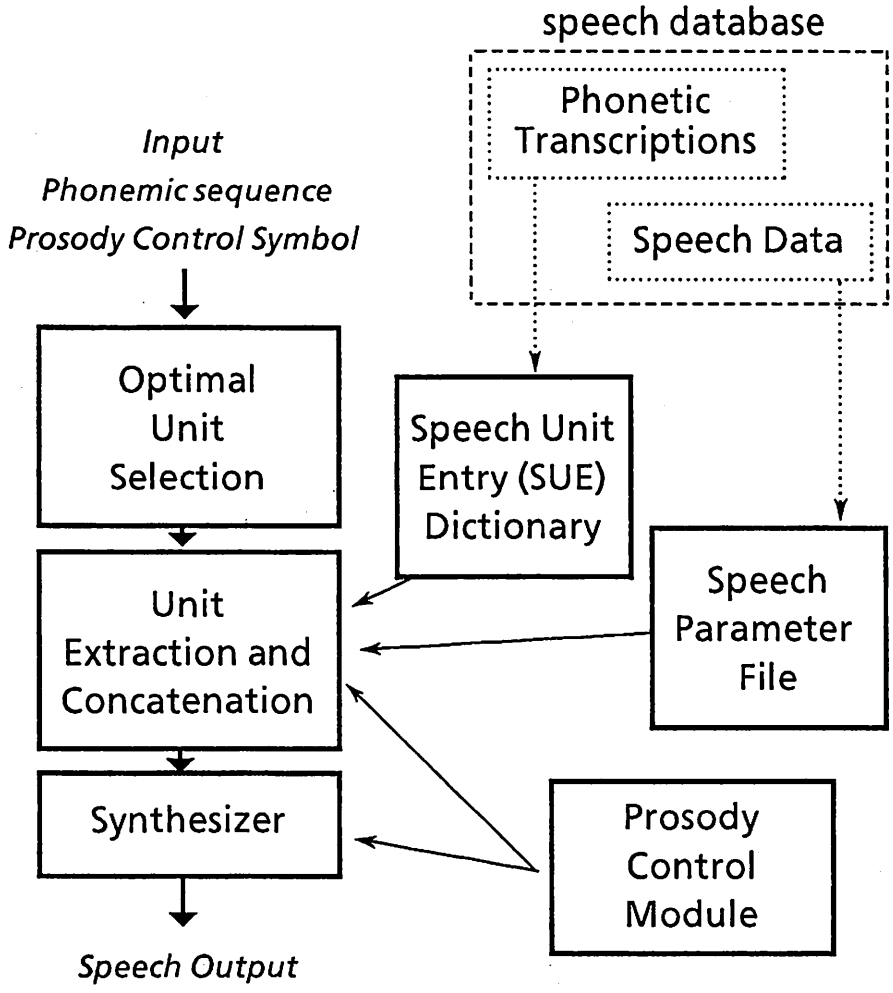

図1　合成システムの構成
Figure 1　Schematic diagram of synthesis system.

3　受聴試験による評価

3.1　試験条件

本稿で用いた合成システムの基本的な性能を評価するために、いくつかの受聴実験をおこなった。試験の内容を表1に、用いたテキストを表2に示す。

3.2　対比較実験1

DP法により選択された単位と、EXP法により選択された単位とを比較し、プリファランスにより評価する実験を行った。実験には、9文からなる模擬電話会話(Aセット)を数文節毎に分割して得られた、17サンプルを用いた(Appendix参照)。被験者は、どちらが『聞きやすい』合成音か判断できるまで、両合成音の対を繰り返し聞き判断の結果を回答した。実験の結果、図2に示すとうりEXP法により選択された単位系列のほうが、より多くの被験者に好まれた。

3.3　対比較実験2

DP法、EXP法により選択された単位と、CV音節単位を用いた合成音とを、対比較により評価する実験を行った。この試験に用いたのは、専門用語からなる文節45文節(Bセット)である。各被験者は、3組の合成音の対(CVとEXP, EXPとDP, DPとEXP)を、各々15文節づつ繰り返し聞き、前節同様『聞きやすい』合成音が判断できた時点でそれを回答した。

試験の結果図3に示すとうり、EXP法により選択された単位と、DP法によるそれとの間にはあまり差が見られなかった。一方、固定した環境から抽出したCVを単位として用いた場合に比べると、どちらの選択手法により選択された単位も80%以上の割合で好まれた。

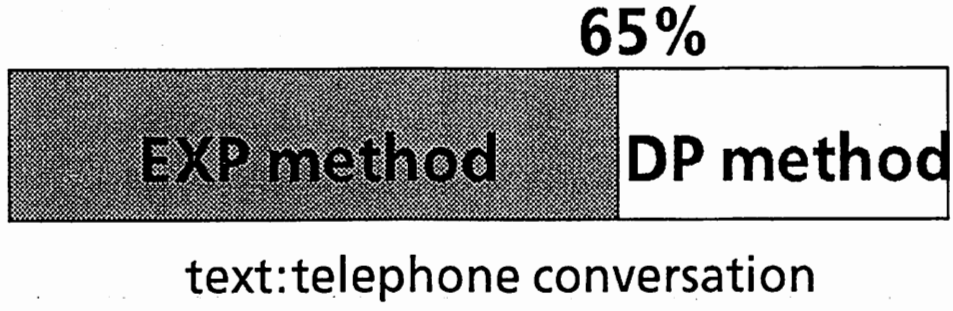

図2　対比較実験1の結果
2つの単位選択法(EXP法とDP法)の比較
Figure2　Subjective preference score of the two selection methods.

表1　受聴試験の内容

試験名	試験内容	評価対象	被験者数	使用テキスト
対比較1	プリファランス	DP法 EXP法	11	Aセット 17(9)
対比較2	プリファランス	DP法 EXP法 CV音節	9	Bセット
単語明瞭度	聞き取り	DP法 EXP法	15	Cセット
音節明瞭度	聞き取り	DP法 EXP法 CV音節 自然発声	12	Bセット

表2　実験に用いたテキスト

セット	内容	サンプル数	例
A	模擬電話会話	9	『通訳電話の国際会議に参加の登録を希望される方は、所定の申込み用紙に、住所氏名と発表聴講の別を明記して、国際会議事務局までお申込み下さい。』
B	専門用語を含む文節	30	『動的計画法と最適化』, 『音節明瞭度の学習効果』
C	頻出する姓	100	『佐藤』,『田中』

3.4　単語明瞭度試験

DP法とEXP法により合成された音声の単語聞き取り試験を、人名(Cセット)を用いて行った。作成した合成音は、頻度順位1位から100位までの、『日本人の姓』[15]で、順位をランダマイズして16名の被験者に3度づつ呈示した。

表3に示す通り、実験の結果得られた了解度は、DP法の94.5%に対し、EXP法では99.0%と高い了解度が得られた。被験者間のばらつきを見ても、EXP法による結果の方が安定している。さらにこれらの結果を、CV-LSP合成による同様の実験結果[1]と比較すると、EXP法で選択された単位により高い単語了解度が得られることが分かる。

3.5　音節明瞭度試験

合成音の品質を音節了解度の観点から評価することを目的に、専門用語からなる文節(Bセット)の書き取り試験を行った。試験は、DP法、EXP法により選択された単位及びCV音節、の3種類の単位による合成音を、各々15文節づつ聞き取ることで行った。被験者は12名である。さらに比較のため、合成音の聞き取り試験の後、自然発声を用いて同様な試験を行った。

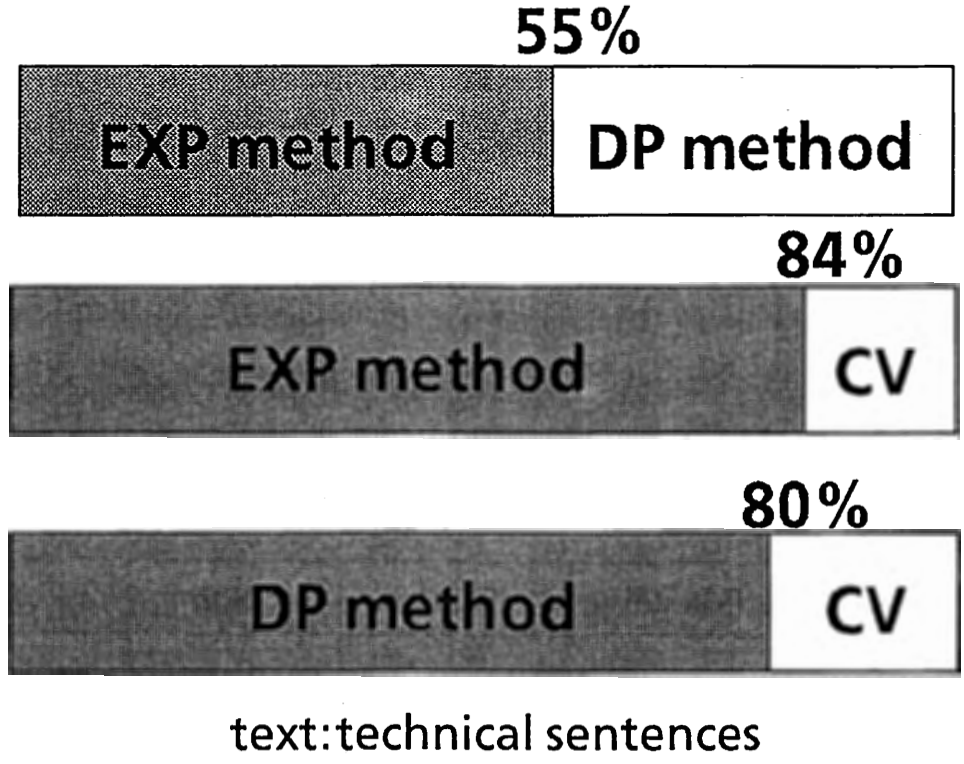

図3　対比較実験2の結果
3つの合成間(EXP法,DP法とCV音節)の比較
Figure3　Subjective comparisons among three types of synthesized speech

表3　選択的な単位使用により得られた単語了解度

選択法	平均了解度	被験者間のばらつき	
		最高	最低
DP法	94.5	100.0	86.0
EXP法	99.0	100.0	96.0

試験結果を、表4に示す。得られた音節了解度は、自然発声で98%、EXP法による単位を用いた合成音で91%、DP法による単位を用いた場合87%、CVを用いた合成音で77%であった。これらから、選択的に単位を使用することでCV音節を用いる場合に比べ高い音節了解度が得られることが分かる。特に、CV法では正解文節との対応が取れない『不明』誤り、2音節以上にわたって脱落が生じる『省略』誤りの起こる割合が高い。(これら誤りの分類は図4に示した。)

さらに、最も了解度の高かったEXP法による単位において音節了解度の低かった(90%未満)音節を、図5に示す。

表4　各手法の音節了解度

	自然発声	EXP法	DP法	CV
音節数	2100	2784	2767	2753
音節誤り	45	201	244	347
省略	2	47	107	262
不明	0	0	19	36
誤り音節数	47	248	370	645
音節了解度	98%	91%	87%	77%

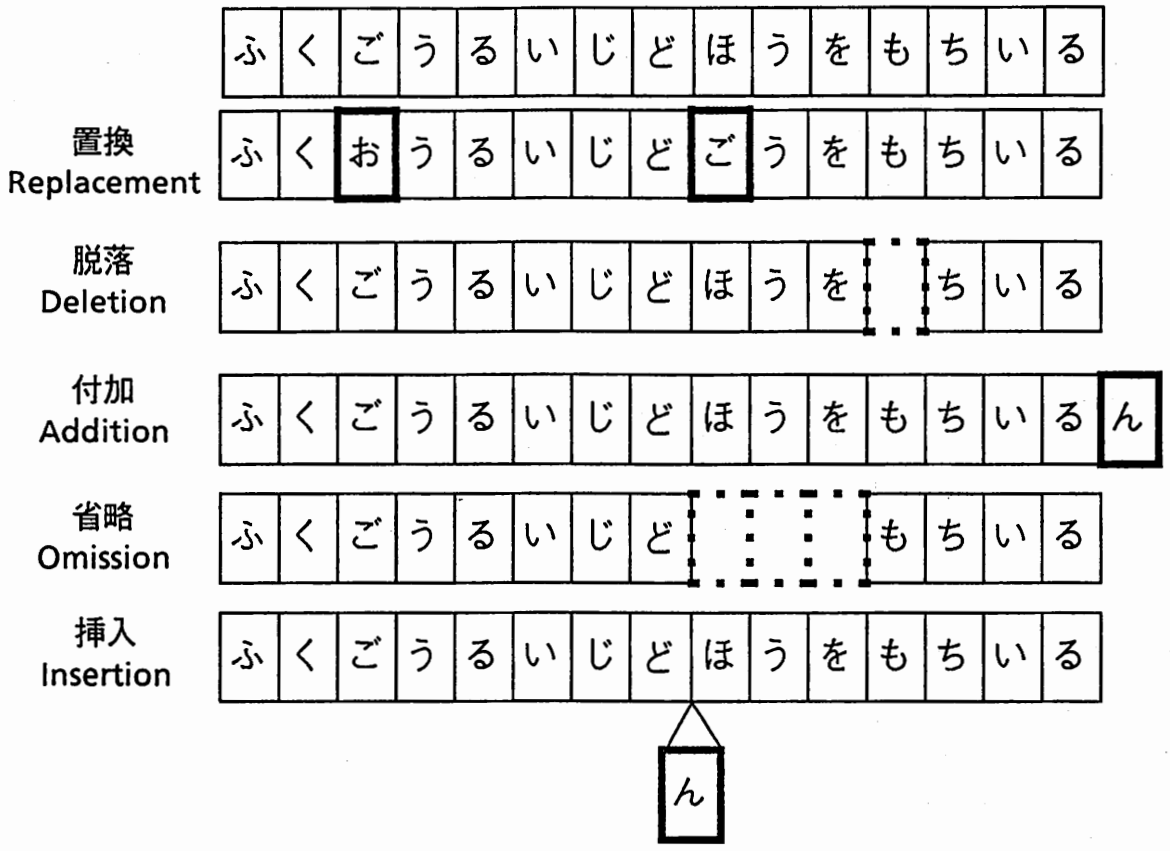

図4　音節了解誤りの分類
Figure 4　Classification of mis-identification

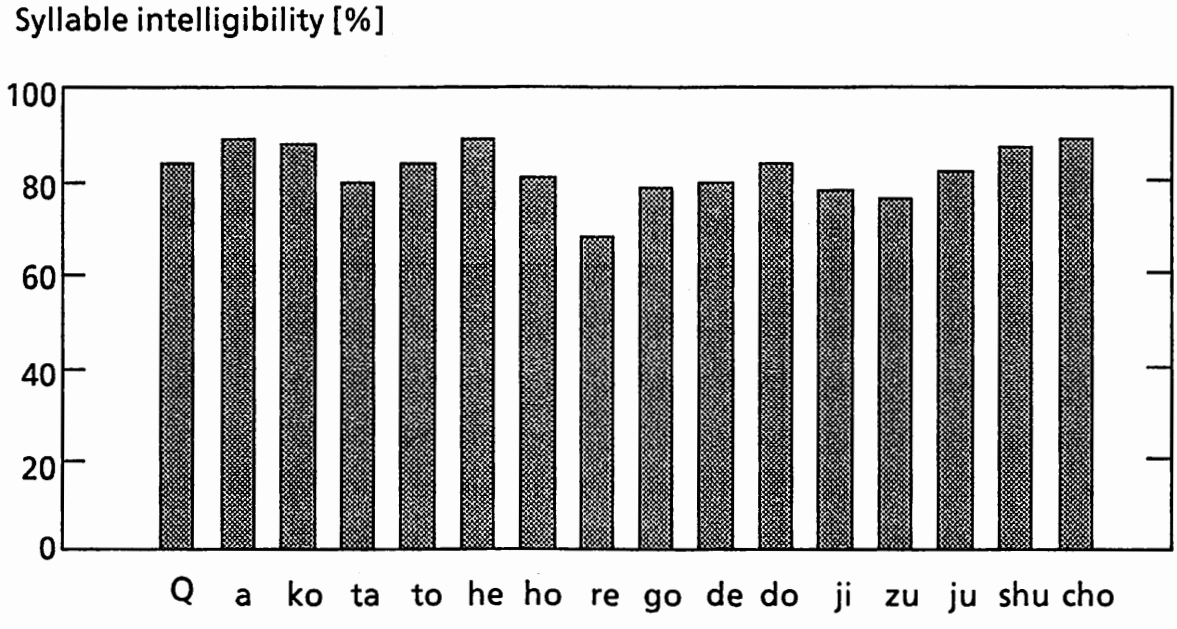

図5　音節明瞭度の低い(90%以下)音節
Figure 5　Syllable intelligibility less than 90%

4. 合成単位の属性と合成音の品質

本節では、音節了解度の試験結果を用いて、合成単位の持つ音響的な特徴と、得られる合成音声の品質との関係を分析する。分析に用いるデータは、3.5節で述べた実験により得られた、専門用語による文節を入力としたEXP法による単位系列とその音節了解度である。分析に用いた単位の総数は1508、平均単位長は3.14(音韻)であった。

4.1 単位長と合成音の品質

用いられた単位の長さ(音節数)と当該単位内の音節了解度との関係を図6に示す。図から、長い単位を用いるほど高い了解度が得られ、特に単音節単位を用いた場合の了解度と、2音節以上の単位を用いた場合の了解度の差が大きいことが分かる。このことから単位長の評価尺度は、単に長さに比例した尺度でなく、『単音節単位の使用を避ける』メカニズムを反映したものが、必要であると考えられる。

4.2 音韻環境の適合度

入力音韻系列に対して、適切な音韻特徴を持つ合成素片を選択するためには、合成単位の抽出環境と使用環境を一致させることが、重要である。本実験では、単位素片の抽出時と使用時における隣接音韻を比較し、両者の音韻的な属性の違いを適合指標として用いた[13]。(例えば、「有声の音韻に後続する素片を抽出し、無声音韻に後続して使用した場合、コストとして10を与える」といったように、この指標が大きいほど環境の適合度は低い。)以下この指標を用いて、音韻環境の適合度と了解誤りとの関係を分析する。

表5に音節了解誤りが生起した単位の適合指標の平均値を、誤りの種類別に示す。どの誤りにおいても、平均適合指標が全単位の平均37.7 (s.d.= 20.1) を上回っている。さらに置換･脱落が起こった単位は、付加･挿入が起こった単位に比べ適合指標が低いことが分かる。生起頻度と合わせて考えると、置換･脱落誤りは、環境が比較的良く適合する単位でも生起し、適合が悪くなるに従い、付加･挿

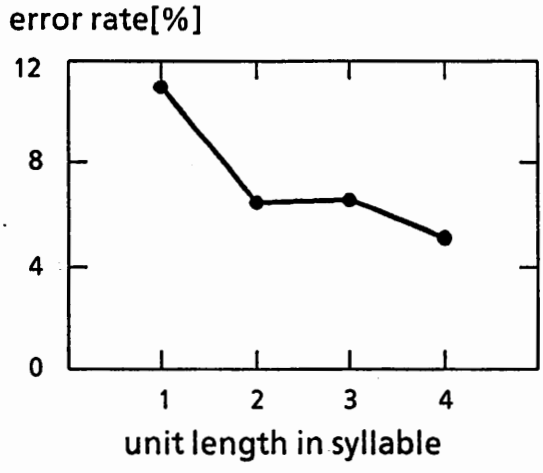

図6 単位長と了解誤り
Figure 6 Syllable intelligibility v.s. unit length

入誤りが生起しやすくなると考えられる。

最も多かった置換誤りを、誤り音節の単位内における位置(単位頭、単位中、単位末、単音節単位)別に分類し、当該単位の適合指標の平均を求めた結果が図7である。適合指標の平均値は、単音節＜単位頭＜単位末＜単位中の順で大きくなっている。即ち、環境が比較的良く適合する単位でも、単位末あるいは単位中に比べ、単位頭で誤りが起きやすいことから、環境類似度の尺度化には、『左側のコンテクストをより重視する』必要があると考えられる。

2音節以上に渡り音節が脱落する省略誤りは、置換誤りに次いで多くの音節誤りの原因となっており、一般に複数の単位にまたがって生じる。これら省略誤りが生起した単位の連鎖を、先頭の単位とそれ以外の単位とに2分して、各々の適合指標を平均した結果が図8である。先頭単位の平均適合指標が 53.5 と非常に高いのに対し、以降の単位の平均値は 39.2 と全体の平均値に近い値になっている。このことから、極度に適合の悪い単位が用いられた場合、音節誤りが以降の単位にも

表5　誤り別適合指標

†省略:連続する2音節以上の脱落
‡括弧内は自然発声における誤りの割合

	誤り数	割合‡[%]	適合指標
置換	146	59(70)	39.7
脱落	41	17(23)	38.9
挿入	7	3(2)	42.6
付加	7	3(0)	49.9
省略†	47	19(4)	-

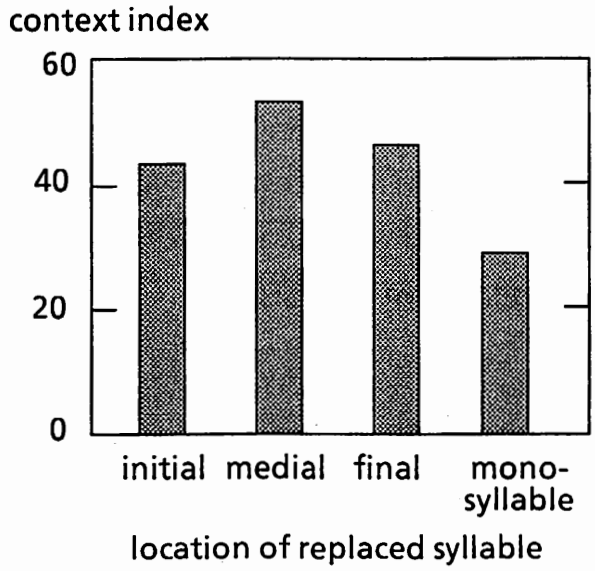

図7　誤り音節の位置と適合指標 (置換誤り)
Figure 7　Mean context index of units includes Replacement error (classified by the location of the replaced syllable)

伝搬しやすくなると考えられる。即ち『適合指標の局所的な集中を避ける』ことも、環境類似度の評価には重要である。

4.3 接続環境

接続環境と了解誤りとの関係を分析するため、用いられた単位を接続音韻環境別に分類し、単位頭音節で誤りが起った割合を図9に示す。誤りの割合から、『接続音韻環境の好ましさは、母音/無声摩擦＜母音定常部＜母音/無声破裂＜母音/有声破裂＜その他、の順である』ことが明らかになった。

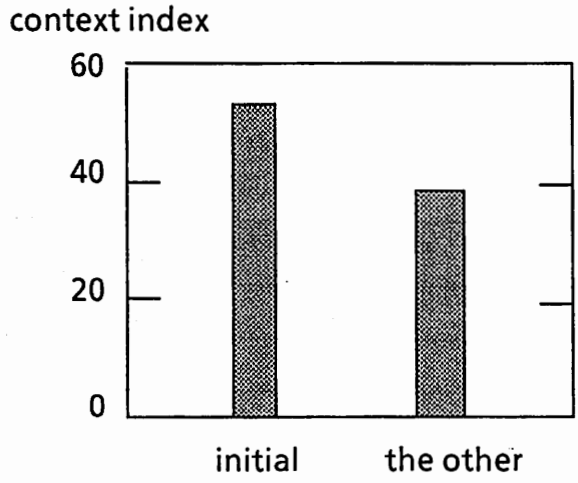

図8 省略誤りを構成する単位の適合指標

Figure 8 Mean context index of units included in Omission errors

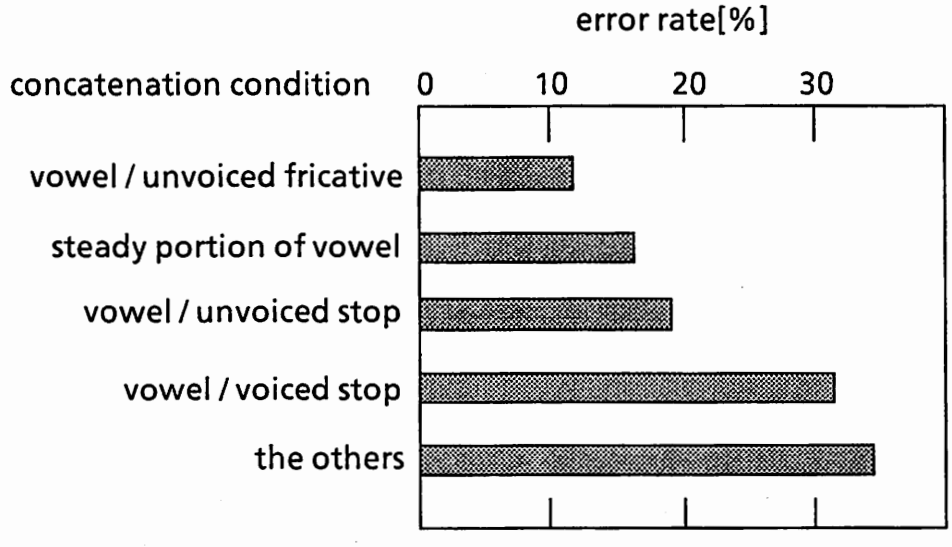

図9 接続環境と了解誤り

Figure 9 Error rate under various concatenation conditions

5 まとめ

単位接続型の規則合成システムにおける最適な単位使用の指針を確立するため、種々の合成単位の組合せから生成された合成音声の品質の評価を行った。先ず、種々の単位を選択的に用いることにより自然性、明瞭性のともに優れた合成音声が得られることを明らかにした。次に、種々の単位により得られた合成音声の音節了解度を分析した結果、単位の性質と音声品質の関係について、以下の知見が得られた。

(1) 合成単位の長さが長くなるに従い、品質が向上する。特に単音節単位と、それより長い単位の間に大きな品質の差が認められた。

(2) 単位の抽出環境と使用環境との差異に着目すると、置換·脱落誤りは、比較的差異が小さな環境でも生じる。さらに差異が大きくなるに従い、付加·挿入誤りが増える。

(3) 了解誤りは、単音節＞単位頭＞単位末＞単位中の順でおこりやすく、抽出環境と使用環境の差異は、左側の環境をより重視して評価すべきである。

(4) 局所的であっても、大きく環境のことなる単位が使われた場合、以降の単位にも影響が及ぶ場合がある。

(5) 接続音韻環境に着目すると、誤りは、母音/無声摩擦＜母音定常部＜母音/無声破裂＜母音/有声破裂＜その他、の順で低く、この順が単位接続点としての適性を反映していると考えられる。

得られた知識を良く反映した単位選択尺度を構成することは、今後の課題である。さらに、これらの尺度を利用することにより、最適な単位素片抽出用音声データベースを設計することも、今後の課題である。

謝辞

研究の機会を与えて下さった、樽松社長に感謝します。また熱心に討論して下さった音声情報処理研究室の諸氏並びに、受聴実験に協力いただいた被験者の皆様に感謝いたします。

文献

[1] 東倉洋一,匂坂芳典:“LSP-CV合成方式とその音声品質”,音響学会音声研資,**S80-47**,pp.371-376 (1980-10)

[2] 斎藤収三,橋本新一郎,脇田寿:“音韻連鎖に着目した音声合成システムについて”,昭52音学会全国大会,1-3-6

[3] 佐藤大和,“CVCと音源要素に基づく(SYMPLE)音声合成”,音響学会音声研資,**S83-69** (1984)

[4] 中島信弥,浜田洋,“音韻環境に基づくクラスタリングによる規則合成法”,信学論(D-II),**J72-D-II**,8,pp. 1174-1179(平1-08)

[5] 岸本直人,東倉洋一:“音声合成用CVファイルにおける音韻環境の考慮”,昭55音響学会全国大会,1-6-20

[6] 野村哲也,佐藤大和:“低次ケプストラム係数の連続性を考慮した音素セグメント接続法”,信学技報,**SP89-65** (1989-11)

[7] 広川智久,箱田和雄,中津良平:“波形編集型規則合成法における波形選択法”,信学技報,**SP89-114**,(1990-01)

[8] 匂坂芳典:“種々の音韻連接単位を用いた日本語音声合成”,信学技報,**SP87-136** (1988-3)

[9] K. Takeda, K. Abe and Y. Sagisaka:“Adaptive manipulation of non-uniform synthesis units using multi-level units transcription”, Proc. Euro.Conf.Speech Tech.(1989-09)

[10] 今井聖:“対数振幅近似(LMA)フィルタ”,信学論(A),**J63-A**,12,pp. 886-893 (昭55-12)

[11] 武田一哉,匂坂芳典,片桐滋,桑原尚夫:“研究用日本語データベースの構築”,音学誌,**44-10**,pp. 747-754(昭63-10)

[12] 今井聖,阿部芳春:“改良ケプストラム法によるスペクトル包絡の抽出”’信学論(A),**J62-A**,4,pp. 473-474(昭54-04)

[13]武田一哉,安倍勝雄,匂坂芳典:“エキスパートシステムを用いた単位選択の検討”信学技報,SP89-113 (1990-1)
[14]安部勝雄,武田一哉,匂坂芳典:“音韻環境に応じた音声合成素片の接続方法の検討”,信学技報,SP89-66 (1989-11)
[15]佐久間:,“日本人の姓”,六芸書房

Appendics

受聴試験に用いたデータ(Aセット)

文節	EXP法を好んだ被験者数
第一回通訳電話国際会議に参加の登録をご希望される方は	7/11
所定の申込み用紙に住所氏名と発表聴講の別を明記して	10/11
国際会議事務局までお申し込み下さい	8/11
はい。こちらは第一回通訳電話国際会議事務局です	11/11
もしもし。通訳電話国際会議への参加を申し込みたいのですけれども	7/11
どのような手続をすればよろしいのでしょうか	10/11
通訳電話の国際会議に参加するためには	11/11
所定の申込み用紙を用いて参加登録することが必要です	11/11
会議に発表するのではなくて聴講するだけだと	4/11
費用はいくらかかりますか	2/11
御発表を希望される場合には	4/11
予稿集代登録料を含めた参加費用は四万円です	8/11
聴講のみの場合は当日の受付も可能で	7/11
予稿集代を含めた費用は三万五千円かかります	6/11
参加登録の申込み用紙はどのようにして手に入れればよろしいのでしょうか	2/11
お名前と御住所をお知らせいただければ	11/11
国際会議の事務局から所定の用紙を送らせていただきます	2/11

Aセットの文章リスト

受聴試験に用いたデータ(Bセット)

帯域圧縮による符号化
電話帯域での音声処理
周波数成分への分解
時間窓のための関数
最適な直交展開系
離散系による余弦展開
共分散法による推定
最尤推定による定式化
偏相関による定式化
変形格子法による実現
複合類似度法を用いる
動的計画法と最適化
統計的な手法を利用する
時間軸の正規化
同期型の構文解析
確率的な句構造文法
誤差逆伝搬による学習法
巡回型の自己相関関数
正規化された直交系列
双一次変換になっている
直交多項式による展開
移動重み係数の意味
声帯音源波形の分析
素性構造の単一化
単一化に基づく言語処理
対論構造の解析手法
導出原理に基づく推論
構成的数学における証明
階層的認識論理の構成
暗黙推論における証明図
変動記述の情報圧縮
線形音素文脈の自動化
混同確率行列の作成
枝かりの閾値関数
統語的情報の利用
文節列選択の目的関数
横型解析に対する適用
格構造表現による述語
音韻列整合度の値
聴感重みづけ誤差の平均値
準最適化法による探索
声道伝達特性の近似
周波数領域での畳込み
音節明瞭度の学習効果

類似音韻列併置による地口