

外国人学習者の日本語音声に対する評価基準について： 韓国人学習者を対象にして

著者	佐藤 友則
出版者	長野県ことばの会
引用	ことばの研究 10: 51-62(1999)
発行年月日	1999-10-10
URL	http://hdl.handle.net/10091/00022426

外国人学習者の日本語音声に対する評価基準について — 韓国人学習者を対象にして —

佐藤友則

1. 研究の目的

外国人学習者に対する音声指導を考えていくうえで、学習者が日本語音声に対しどのような評価基準を持って評価を行っているかを知るとは、非常に重要である。誤った評価基準を持ち、自然ではない日本語音声を自然だと評価しては、学習者の自然な日本語音声把握が困難になると予想されるからである。それは、学習者が日本語音声を習得する過程で大きな妨げになろう。また日本国内以外の環境で日本語を学習する場合、自然な日本語に触れる機会が少ないため、日本語音声に対する正しい評価基準が作成されにくいと考えられる。そこで本研究では、より自然な日本語音声習得のための基礎研究として、韓国の大学で日本語を学習した学習者達を対象に合成音声を用いて聴取実験を行い、佐藤(1995)で得た日本人被験者の評価基準と韓国人学習者の評価基準との比較を行う。さらに、この韓国人学習者達を対象に、約1ヵ月間の日本国内での日本語学習の後に再度実験を行い、彼らの評価基準に何らかの変化が生じたかを確認する。

佐藤(1995)では、日本人が日本語音声の評価する際、高さ・長さ・強さの3つの韻律的要素のうち、どの影響力が最も大きいかをみるために合成音声を用いて聴取実験を行い、高さが最も大きな影響力を持っていることを明らかにした。本研究でも同じ音声を用いて聴取実験を行う。

2. 研究の方法

2-1. 実験文の作成

外国人学習者の音声と日本人の音声をもとに合成音声を作成するため、イントネーション・プロミネンス・特殊拍・学習者にとって発音が

困難な単音などの音声要素を考慮し、6つの実験文を作成した。

2-2. 実験音声の選定

この実験文をインフォーマントに読ませて録音した。インフォーマントは、初中級の韓国語話者1名(ソウル出身)と中級の中国語話者1名(長春出身)、日本人の東京語話者1名である。次に、日本人30名に韓国語話者と中国語話者の音声を聴かせ、各々6つの音声の中から日本語として最も不自然に感じられる音声を1つずつ選ばせた。その結果、「絶対おいしいって店なんだけど、はいる？」という文が最も不自然と評価されたので、この文を用いて合成音声を作成することにした。

2-3. 音声の分析

合成の際に必要なデータを収集するため、韻律的要素に焦点を当ててインフォーマントの音声を分析した。分析には、音声分析機 CSL(注1)と音声分析ソフト「音声録聞見 + β 」(注2)を用いた。まず高さを「音声録聞見 + β 」で分析しピッチ曲線を得た。以下はそれに基づく分析である。

- a. 「ぜったいおいしい」の[a]から[o]の前半まで、中国語話者・韓国語話者とも Fo 値の下降が急激にあるが、東京語話者の下降の幅は小さい。
 - b. 中国語話者の「おいしい」の[i]は、始めから終わりまで Fo 値の変化がないが、他の2者は高く始まり終末部は大きく下がっている。
 - c. 「はいる?」の韓国語話者の Fo 値は、[a]の終末部から文末まで上がり続けているが、東京語話者の Fo 値のピークは[a]の終末部にある。
- 次に CSL で長さを分析し、1万分の1秒単位のデータを得た。

- a. 韓国語話者は「おいしい」の[o]を、藤崎・杉藤(1977: 63)が長母音の判断境界とした長さ(注3)に近い長さで発話している。
- b. 中国語話者の「はいる?」の[ω]は、東京語話者のそれより 0.0421 秒長い。
- c. 韓国語話者は「おいしいって」と「店」の間に不要なポーズを入れている。

2-4. 合成音声の作成

分析で得たデータをもとに、高さ・長さ・強さを様々に入れ換えて合成音声を作成した。韓国語話者の語頭の「ぜ」[dze]など問題がある単音も散見されたが、今回は単音には手を加えなかった。合成音声は、外国人

学習者の音声と日本人の音声の組み合わせにより以下の4組からなる。

中日: 中国語話者の単音+日本人の韻律的要素

韓日: 韓国語話者の単音+日本人の韻律的要素

日中: 日本人の単音+中国語話者の韻律

日韓: 日本人の単音+韓国語話者の韻律

また、各組は、高さ・長さ・強さの組み合わせにより7つの合成音声からなっているので、合成音声の数は4組×7で計28になる。

3. 韓国人学習者に対する聴取実験

3-1. 実験の目的

この実験の目的は、佐藤(1995)で得た日本人の日本語音声に対する評価基準と韓国人学習者の評価基準を比較し、韓国人学習者が日本人と同様の評価基準を持っているか、また1ヵ月間の日本での日本語学習を経て、その評価基準に何らかの変化が生じたかを確認することである。

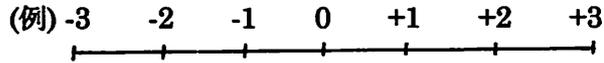
3-2. 実験の方法

実験は、韓国・全羅北道の中級の学習者20名が東北大学に来て受講した日本語コースを利用して行われた。このコースは、1994年の7月半ばから8月半ばまでの約1ヵ月間、週に20時間の日本語の授業を東北大学の大学院生が担当するという内容であった。そこで、韓国人学習者が来日してすぐと、韓国へ帰国する直前の2回、聴取実験を行った。

実験は、インフォーマントのオリジナル音声と、韻律的要素を入れ換えた合成音声をペアにして聴かせ、オリジナル音声と比較して合成音声が日本語として自然になったか不自然になったか、その差異を評価させる方法で行った。オリジナル音声との差異が大きいと評価された合成音声ではどの韻律的要素を入れ換えたかをみることにより、各韻律的要素が評価に与える影響力を知ることができる。

評価尺度は7段階で、合成音声がオリジナル音声と変わらないと評価すれば0、自然になったと評価すれば+1,+2,+3のいずれか、不自然になったと評価すれば-1,-2,-3のいずれかを評価シートにチェックするよ

う指示した。以下に評価シートの1例を挙げる。



また、オリジナル音声と合成音声のペアからなる28個の音刺激を聞かせる前に、評価の方法に慣れさせ評価基準を作らせる目的で、4つの合成音声の組からそれぞれ1個の音刺激をランダムに選んでダミーとして聞かせた。よって被験者は計32個の音刺激を評価したことになる。

さらに、合成音声に対する耳慣らしのために、本実験を始める前にオリジナル音声と合成音声のペアからなるウォーム・アップを2問聴かせた。ウォーム・アップには「コーヒー飲まない？」という文を用いた。

3-3. 実験1 (来日直後の実験) の結果

来日直後の韓国人被験者に対する実験結果を処理して得た、合計値・平均値・標準偏差を[表1]にあげる。また佐藤(1995)で得た、日本人被験者64人に対する実験結果を[表2]にあげる。

[表1] 実験1 (来日直後の韓国人被験者による実験) の結果 N=20

合成音声	計	平均値	標準偏差	合成音声	計	平均値	標準偏差
A 中日一高	9	0.45	1.43	0 日中一高	-14	-0.70	1.10
B 中日一長	6	0.30	0.78	P 日中一長	2	0.10	1.22
C 中日一強	8	0.40	0.92	Q 日中一強	1	0.50	1.25
D 中日一高長	-2	-0.10	1.70	R 日中一高長	-4	-0.20	1.69
E 中日一高強	13	0.65	1.467	S 日中一高強	-15	-0.75	1.41
F 中日一長強	0	0	0.77	T 日中一長強	2	0.10	1.14
G 中日一高長強	13	0.65	1.53	U 日中一高長強	-12	-0.60	1.07
H 韓日一高	14	0.70	1.90	V 日韓一高	-21	-1.05	1.77
I 韓日一長	-9	-0.45	1.02	W 日韓一長	1	0.05	1.07
J 韓日一強	8	0.40	0.80	X 日韓一強	12	0.60	1.07
K 韓日一高長	18	0.90	1.81	Y 日韓一高長	-35	-1.75	1.37
L 韓日一高強	13	0.65	1.68	Z 日韓一高強	-17	-0.85	1.59
M 韓日一長強	1	0.05	0.81	AA 日韓一長強	-1	-0.05	1.28
N 韓日一高長強	9	0.45	1.94	AB 日韓一高長強	-24	-1.20	1.60

[表 2] 日本人被験者による実験の結果

N=64

合成音声	計	平均値	標準偏差	合成音声	計	平均値	標準偏差
A 中日一高	77	1.20	0.86	O 日中一高	-69	-1.08	0.60
B 中日一長	1	0.02	0.42	P 日中一長	-17	-0.27	0.62
C 中日一強	0	0	0.36	Q 日中一強	-1	-0.02	0.38
D 中日一高長	68	1.06	1.13	R 日中一高長	-89	-1.39	0.66
E 中日一高強	93	1.45	0.92	S 日中一高強	-85	-1.33	0.54
F 中日一長強	1	0.02	0.33	T 日中一長強	-11	-0.17	0.52
G 中日一高長強	91	1.42	0.83	U 日中一高長強	-91	-1.42	0.59
H 韓日一高	80	1.25	0.89	V 日韓一高	-128	-2.00	0.62
I 韓日一長	-3	-0.05	0.36	W 日韓一長	-27	-0.42	0.71
J 韓日一強	2	0.03	0.31	X 日韓一強	-3	-0.05	0.45
K 韓日一高長	78	1.22	0.93	Y 日韓一高長	-153	-2.39	0.63
L 韓日一高強	68	1.06	0.75	Z 日韓一高強	-132	-2.06	0.66
M 韓日一長強	-1	-0.02	0.42	AA 日韓一長強	-52	-0.81	0.59
N 韓日一高長強	99	1.55	0.97	AB 日韓一高長強	-132	-2.06	0.75

[表 1]をみると、日本人の結果である[表 2]に比べて韓国人被験者の標準偏差が大きいことがわかる。[表 2]では、標準偏差が 1 を越える合成音声は D:中日一高長の 1 つだけであるが、[表 1]では、標準偏差が 1 を越える合成音声が 23 もある。このように標準偏差が大きいということより、個々の韓国人被験者の評価基準に、日本人にないバラつきがあることが予想される。そこで D-Score(注 4)を求め個々の被験者のデータと日本人の平均との差をみた。日本人の平均との差がなければ D-Score は 0 になる。

[表 3]韓国人被験者の個人データと日本人の平均値の D-Score

被験者	k1	k2	k3	k4	k5	k6	k7	k8	k9	k10
D-score	0.25	0.23	0.38	0.36	0.25	0.20	0.36	0.32	0.30	0.27
被験者	k11	k12	k13	k14	k15	k16	k17	k18	k19	k20
D-score	0.24	0.24	0.29	0.35	0.29	0.27	0.22	0.38	0.32	0.20

[表 3]より、個々の韓国人被験者が日本人の平均と異なる評価をしていることが分かる。また韓国人被験者が各合成音声毎にどのような評価を下しているかをみるために、[表 4]に評価のカテゴリ一度数をあげる。

[表 4] 韓国人被験者による評価のカテゴリ一度数

合成音声	-3	-2	-1	0	1	2	3	合成音声	-3	-2	-1	0	1	2	3
A 中日一高	0	2	4	3	7	2	2	O 日中一高	0	5	9	1	5	0	0
B 中日一長	0	0	2	12	4	2	0	P 日中一長	0	1	7	4	6	1	1
C 中日一強	0	0	2	12	2	4	0	Q 日中一強	0	1	4	4	7	3	1
D 中日一高長	2	3	4	1	7	2	1	R 日中一高長	1	5	4	2	4	3	1
E 中日一高強	0	1	5	2	7	2	3	S 日中一高強	1	5	9	0	4	0	1
F 中日一長強	0	0	5	11	3	1	0	T 日中一長強	0	1	5	8	4	1	1
G 中日一高長強	0	2	4	1	8	2	3	U 日中一高長強	0	4	9	2	5	0	0
H 韓日一高	0	4	4	0	2	6	4	V 日韓一高	4	8	2	0	3	3	0
I 韓日一長	1	2	5	9	3	0	0	W 日韓一長	0	1	6	6	5	2	0
J 韓日一強	0	0	1	13	3	3	0	X 日韓一強	0	0	3	7	6	3	1
K 韓日一高長	0	4	2	0	4	6	4	Y 日韓一高長	5	11	2	0	0	2	0
L 韓日一高強	0	4	2	1	5	6	2	Z 日韓一高強	2	5	9	0	1	2	1
M 韓日一長強	0	1	2	13	3	1	0	AA 日韓一長強	1	0	8	4	4	3	0
N 韓日一高長強	0	5	4	0	3	4	4	AB 日韓一高長強	2	11	3	0	2	1	1

この[表 4]をみると、全体の約 60%に当たる 17 の合成音声(A,D,E,G, H,K,L,N,O,P,R,S,U,V,Y,Z,AB) で、オリジナル音声と同一という評価は少ないがプラスとマイナスの評価が多い双峰形の分布をなしていることが分かる。そして双峰形をなした合成音声のうち、Pを除いた全ての合成音声で高さの入れ換えがなされている。このことは、韓国人被験者は、評価に大きな影響を与える高さを入れ換えた合成音声において、オリジナル音声と合成音声が同一ではないと判断することはできるが、プラスもしくはマイナスにどの程度変化したかという評価基準はあいまいであることを意味する。

以上により韓国人被験者の評価基準は、個々の被験者ごとにバラつきが大きく、かつ日本人の評価基準よりもあいまいなものであることが分

かる。この評価基準のバラつきは、被験者の学習環境によるところが大きいと考える。ほとんどの被験者が韓国国内でのみ日本語を学習しており、自然な日本語音声を大量にかつ長時間聴くという環境にはない。そのため、どのような音声が日本語として自然かという評価基準があいまいになっており、今回の実験でその影響が現れたと言える。今後、日本語以外の言語を母語とする日本語教師の養成を考えていくうえでは、彼らに日本語音声の正確な評価基準を身につけさせるために、教師養成の段階から大量に日本語の自然な音声を聴かせるなどの指導が必要となるだろう。

3-4. 実験2（帰国直前の実験）の結果

韓国人被験者が帰国する直前、つまり来日してから約1ヵ月経過した時期に実験2を行った。実験の方法は実験1と同様である。

実験の結果を分析したところ、実験1よりは標準偏差が小さくなったが、日本人の結果より大きかった。さらにカテゴリ一度数をみると実験1同様に双峰形の分布をなしていたので、評価にバラつきがあることがわかった。そこで今回もD-Scoreを求め、日本人の平均との差をみた。

[表5]に、実験1と実験2のD-Scoreをあげる。

[表5]実験1と実験2のD-Score（日本人の平均との差）

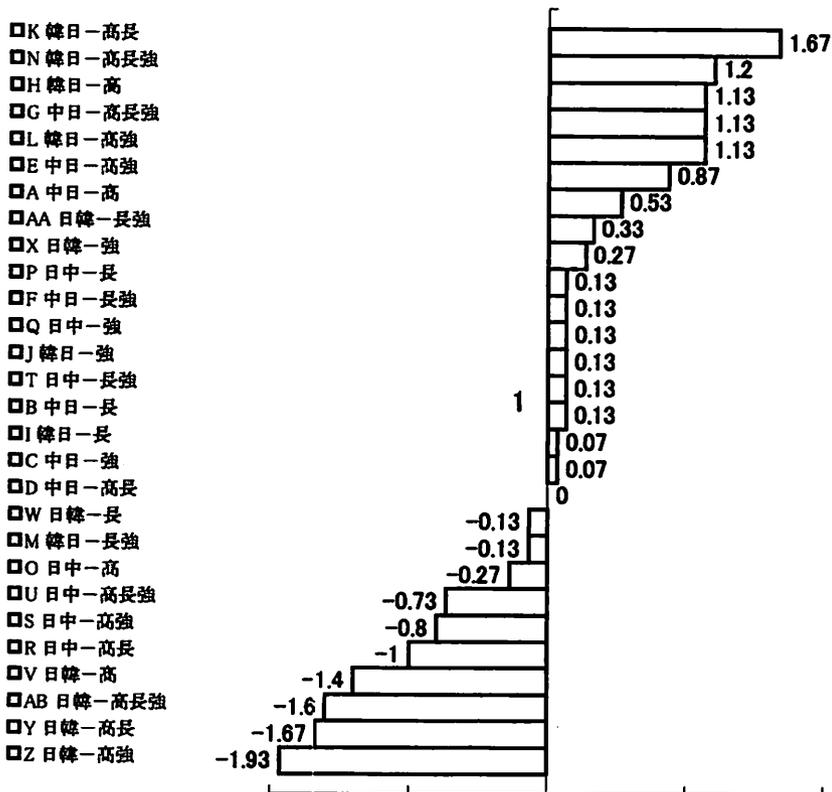
被験者	k1	k2	k3*	k4*	k5*	k6*	k7*	k8*	k9*	k10
実験1	0.25	0.23	0.38	0.36	0.25	0.20	0.36	0.32	0.30	0.27
実験2	0.31	0.29	0.30	0.29	0.10	0.17	0.29	0.20	0.27	0.40
被験者	k11	k12	k13*	k14*	k15*	k16*	k17*	k18*	k19*	k20*
実験1	0.24	0.24	0.28	0.35	0.29	0.27	0.22	0.38	0.32	0.20
実験2	0.34	0.33	0.16	0.29	0.22	0.14	0.21	0.22	0.24	0.12

*印は、実験1よりも実験2が日本人の平均に近くなった被験者

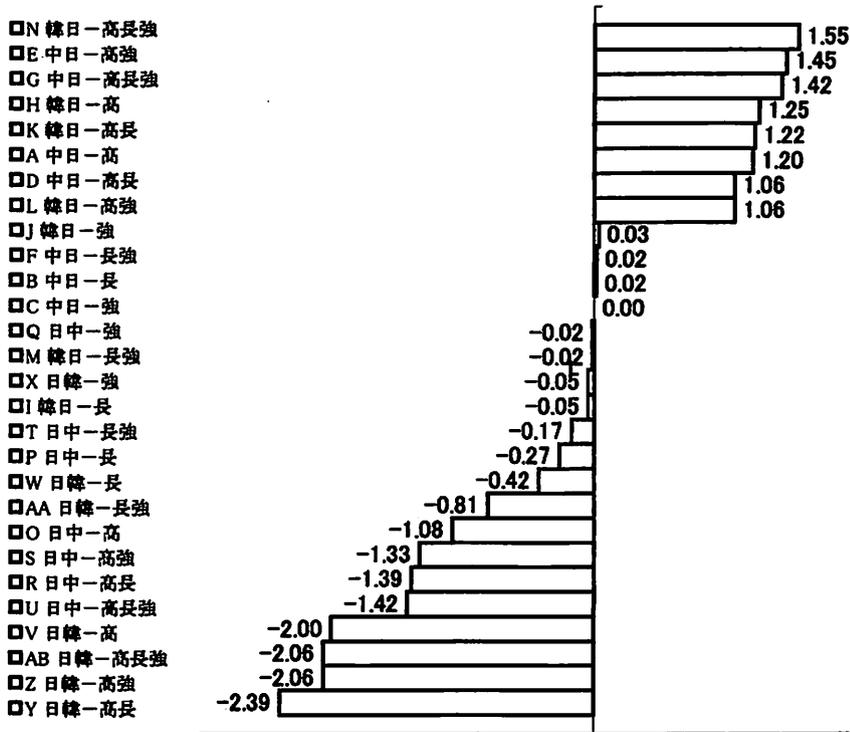
[表5]より、実験1ではD-Scoreが0.1台の被験者は1名だけだったのが実験2では6名に増えており、日本人の平均に近づいていることが分かる。また20名の被験者のうち15名([表6]の*)が、実験1より日本人の平均に近い評価をするようになっている。この15名に共通して見

られる特徴は、1ヵ月間の実習を通し発音・聴解に進歩が見られたことである。来日当初はコミュニケーションに支障をきたすほど発音・聴解に問題が見られたが、1ヶ月後にはそれらの問題が非常に少なくなり、文法の問題はあれど十分コミュニケーションできるようになった。一方日本人の平均から外れた評価をするようになった5名の被験者には、実習の終わりまで発音・聴解に大きな改善が見られなかった。

次に、実験2で日本人に近い評価をするようになった15名の被験者の平均値によるグラフを作成し、日本人64名の平均値によるグラフとの比較を行った。[図1]に15名の被験者の平均値によるグラフを、[図2]に64名の日本人によるグラフをあげる。



[図1] 15名の韓国人被験者の平均値によるグラフ N=15



[図 2] 64名の日本人被験者の平均値によるグラフ N=64

この2種類の数値について、順位相関をスピアマンの計算方法で求めたところ、 $\gamma_s = 0.868$ という高い相関がみられる結果を得た。よって[図1]の15名の韓国人被験者は、日本人と近い評価をしたといえる。つまりこの15名の韓国人被験者も、日本人被験者と同様に、高さの影響力が最も大きく長さ・強さがそれに次ぐと評価したわけである。

しかし今回の結果が日本人の評価結果と異なる点は、中日・日中などの中国語話者と日本人との合成音声、オリジナル音声とあまり変わらないと評価されている点である。日本人の評価結果ではオリジナルの音声と大きく異なると評価されていたE(中日-高強: 平均値 1.4531)が、韓国人被験者の評価では平均値 0.8667となり、それほど大きな差異があるとは評価されていない。他にも、O(日中-高)やD(中日-高強)など、オリジナル音声との差が日本人の評価よりも小さくなっている合

など、オリジナル音声との差が日本人の評価よりも小さくなっている合成音声は、いずれも中国語話者と日本人との合成音声である。一方、日韓・韓日といった韓国語話者と日本人の合成音声は、高さを入れ換えた合成音声が上位3位、下位4位を占めるなど、オリジナル音声と大きく異なると評価されている。このことは、被験者が韓国語話者であることと何らかの関連があるのではないかとと思われる。

上記の、日本人に近い評価をするようになった15名の被験者達は、実習の期間中に日本人学生とよく一緒に行動するなど、日本語に積極的に接する努力をしていた。そのような被験者達が、来日直後の実験では評価にバラつきがみられたのに、約1ヵ月後の実験では日本人に似た評価を行ったことは注目し得る。日本国内という目標言語環境内で自然な日本語を多く聴いたことが、学習者の日本語音声に対する評価基準作成に大きな役割を果たしたと考える。またこの学習者達は、教室外の環境でも、遠足やドライブなど日本人と自由に接触する機会が多くあったため、教室内の正確で細かい聴き取りのみでなく、コミュニケーションに必要とされる大まかな聴き取りにも注意を向けることが多かった。そのことも評価基準作成の一助になったと考えられる。そのような環境にない学習者には、日本人が聴き取りの際に重要な判断材料としている音声要素を今後の研究で明らかにしてゆき、それを適切に指導していくことが望まれる。

今回の結果により、日本語音声を自然な環境で大量に聴くことは、日本語音声に対する評価基準を作らせるためにある程度有用であると言える。このことは裏返せば、国外で日本語を指導する際、教師は意識して学習者に多くの自然な日本語音声を聴かせ、正しい評価基準作成の手助けをする必要があることを意味する。自然な日本語音声を習得した教師が早い段階から日本語を使って授業する、自然な録音のテープを効果的に利用する、ネイティブの日本語話者との会話の時間を作る等の工夫が必要であろう。

4. まとめと今後の課題

この研究では、外国人学習者の日本語音声に対する評価基準はどのようなものか、またその評価基準は日本での日本語学習により変化したかどうかを明らかにしようとした。そして韓国人学習者に合成音声を用いて聴取実験を行い、以下の結果を得た。

- ①韓国人学習者は、自然な日本語音声にあまり多く触れていない段階では、日本語音声に対する評価基準があいまいであった。
- ②約1ヵ月間日本で日本語を学習した韓国人学習者のうち3/4の学習者は、日本人と同様の評価をした。

この研究では、高さ・長さ・強さの3つの韻律的要素を個々に独立したものとして影響力の比較を行ったが、これらが干渉し合って評価に影響を与えていることも考えられる。またこの研究では、中級レベルの中国語話者および韓国語話者を外国人インフォーマントとして用いて合成音声を作成したが、これを様々な母語の中級以外の学習レベルの学習者に変えることにより、今回とは異なった結果が得られる可能性もある。さらに、正しい日本語音声に対する評価基準が、自然な日本語音声の生成にどのように関与しているかを明らかにしていかなければならない。これらを今後の課題とし、さらに視野を広げて研究を深めていきたい。

注

- (1)CSL: KAY Elemtric 社の音声分析機。(Computerized Speech Lab)
- (2)音声録聞見+ β : 音声分析ソフト「音声録聞見+ β . PARCOR 分析再合成」の略。今川・桐谷(1989)参考。
- (3) 藤崎・杉藤が長母音の判断境界とした長さ: 短文においての判断境界は 0.168 秒、単語においての判断境界は 0.156 秒としている。
- (4)D-Score: 個別データと平均との差を示す数値尺度。

参 考 文 献

- 今川博・桐谷滋(1989)「DSPを用いたピッチ・フォルマント実時間抽出とその発音訓練への応用」『電子情報通信学会技術報告』SP89-36.
- 大山玄・三浦一郎(1990)「日本語学習者のプロソディーに関する研究」『日本語音声における韻律的特徴の実態とその教育に関する総合的研究 研究報告3』 pp.98-101.
- 佐藤友則(1995)「単音と韻律が日本語音声の評価に与える影響力の比較」『世界の日本語教育』第5号 pp.139-154.
- 杉藤美代子(1989)「現代の日本語音声研究の課題」『日本語学』第8巻 第3号 明治書院 pp.4-13.
- 土岐哲・村田水恵(1989)「イントネーション」『外国人のための日本語例文・問題シリーズ 発音・聴解』荒竹出版 pp.37-55.