（信州大学留学生センター紀要第5号　2004年3月）

ポーズと長さが音声評価に与える影響力の比較

—外国人学習者の日本語音声評価において—

佐　藤　友　則

キーワード：ポーズ・長さ・合成音声・評価への影響力・自己モニター

要旨

本研究では、音声指導上重要なシラバスである「ポーズ」と「音の長さ」に焦点を当てた。そして、日本人が外国人の音声を評価する際に、「ポーズ」と「音の長さ」のどちらが評価に与える影響力が大きいかを比較した。そのため、2つの音声項目を入れかえた合成音声を作成して85名の日本人を対象に聴取実験を行ったところ、「ポーズ」の影響力のほうが「音の長さ」の影響力よりも大きいという結果を得た。この結果は、「ポーズ」指導の重要性を示唆するものであり、今後の音声指導を考えていくうえで意味がある結果だと言える。本稿では、最後にポーズの指導法についても言及した。

1．研究の目的

昨今、外国人学習者への音声指導に関する研究が積極的に行われている。特に、音の長短の問題は、古くから研究が進められ、多くの知見が得られてきている分野である。外国人学習者にとって、長母音の認識および生成は非常に困難かつ重要な問題であり、長母音指導の重要性については今更強調する必要はない。しかし、長さ同様、音の持続時間が関係するポーズの指導の重要性についてはどうであろうか。

本稿では、ポーズを「話しことばの途中におこる声や息の切れ目」のことであると定義する。ポーズは、それを効果的に挿入することにより、強調、分かりやすさ、スムーズさ等に非常に大きな影響を与えるものである。話し手が自らの意図を聞き手に正確に伝達するためには、適切なポーズが作られなければならない。また、不適切なポーズが多く入っている音声は、スムーズさが失われ、非常に聞きづらくなる。よって、初級学習者に関しては、不適切なポーズを挿入させない指導が必要であり、中上級学習者に関しては適切なポーズ挿入の指導が必要とされる。

しかし、現在まで行われてきた音声指導では、高さや長さ同様に、ポーズをも重視して指導してきたと言えるであろうか。音声に重点をおいて定期的に指導しているという日本語教育機関であっても、ポーズに焦点を当てた指導が行われているところは多くないと予想される。しかし、もし日本人がポーズをも重視して外国人学習者の音声を評価している

のであれば、ポーズも、より重点をおいて指導されなければならない。

そこで、本研究では、合成音声を用いて日本人を対象に聴取実験を行い、日本人が外国人の日本語音声を評価する際、長さとポーズのどちらがその評価に大きな影響を与えるかを検証することにする。

2．研究の方法

2－1．インフォーマント

インフォーマントには、スペイン語を母語とする外国人日本語学習者と東京出身の日本人の2名を用いた。

スペイン語を母語とするインフォーマントの音声採取時点での年齢は26歳であった。この学習者は、日本語に関しては全く未習の状態で来日し、音声採取段階での日本語学習期間は1年半であった。しかし、短期間であっても会話能力の習得が早く、会話能力に関しては中級といっても過言ではないレベルに達していた。

本研究のインフォーマントに、欧米言語の一つであるスペイン語の話者を用いた理由は、音節構造が日本語と大きく異なり、拍の等時性の理解および特殊拍・長母音の認識・理解が困難な学習者が多いためである。この学習者の発話にも、音声上の諸問題が多く存在していた。また、本研究では自然な自由発話を採取することも重要な方針であったため、本研究者とリラックスして話すことができるインフォーマントが必要であった。このインフォーマントは、本研究者が1年半日本語を指導した学習者であり、緊張することなく自然に会話ができる関係にあった。

この外国人インフォーマントに加え、長さとポーズの合成音声作成に必要な日本語音声話者として東京出身の男性に依頼し、日本人インフォーマントとした。この日本人インフォーマントの音声採取時点での年齢は21歳であった。

2－2．実験音声の採取

今回の実験では、学習者のより自然な状態の発話を利用したいと考え、統制された実験文を読み上げさせて採取するのではなく、自由発話を採取することにした。実験文を読み上げさせると、研究者が意図した通りの音声を表現させられるというメリットはあるが、文字を読むという作業が加わることによる不自然さ、自分の自由発話ではないことによる不自然さというデメリットも存在する。

そこで、2001年3月に、雑音が入らず人の出入りがない部屋において外国人インフォーマントと本研究者が雑談し、リラックスした状態での自由発話を全て録音した。雑談の内容は、この外国人インフォーマントの今後の進路など多岐にわたった。録音時間は2時間半ほどであった。録音にはSONYのDAT TCD-D8を用いた。

録音された発話音声の中から、以下の2点に注意して実験音声を選択した。

①本研究者のあいづち、発話などが少ししか入っていない。

②短文レベルで音声評価を行うための十分な長さ(20秒程度)がある。

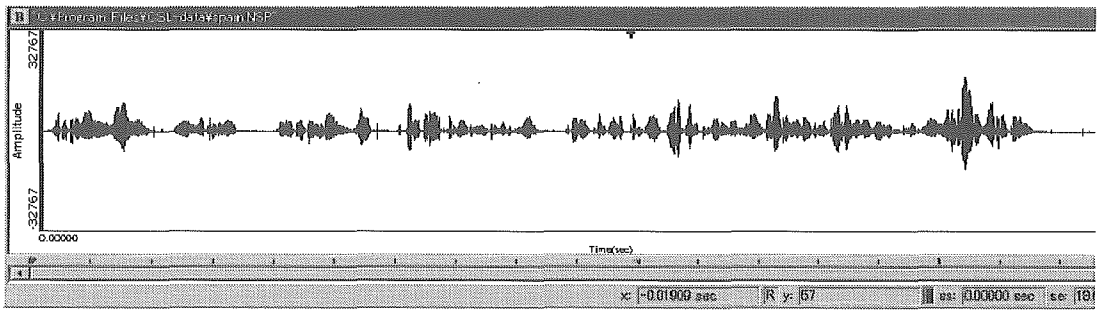
上のような基準により選択された外国人インフォーマントの音声は、

「八十年前に、私のおじいさんはイタリアから外国に行きましたね。今日は反対、私はヨーロッパに帰りますね。同じ問題、いつも経済の問題ですね。面白いですね、ほんとに。大変です。」

である。実際には、上述したようなスムーズな発話内容ではなく、以下述べていくように音声上の問題を多く含んだ音声である。

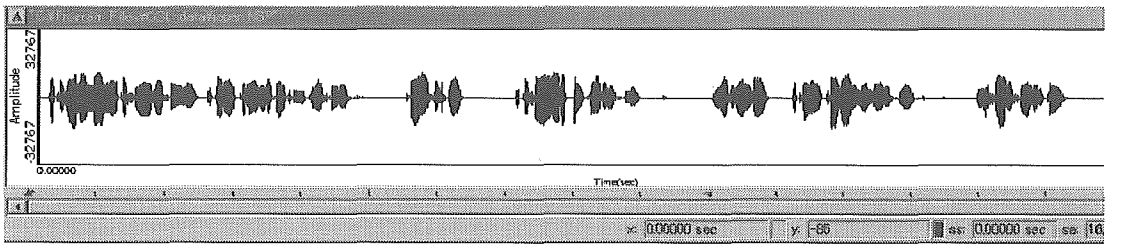
2－3．実験音声の分析

以下に、外国人インフォーマント音声のウェーブフォームを[図１]にあげる。このデータは、採取音声をDATから音声分析器CSL4400に入力後、表示したものである。



[図１] 外国人インフォーマント音声のウェーブフォーム

次に、日本人インフォーマント音声のウェーブフォームを[図2]にあげる。



[図２] 日本人インフォーマント音声のウェーブフォーム

分析結果について述べる前に、参照しやすいように実験文をひらがな表記にして以下にあげる。

「はちじゅうねんまえに　わたしの　おじいさんは　イタリアから　がいこくに　いきましたね。きょうは　はんたい　わたしは　ヨーロッパに　かえりますね。おなじ　もんだい　いつも　けいざいの　　もんだい　ですね。おもしろいですね

ほんとに。たいへんです。」

2－3－1．長さの分析・比較

まず、外国人インフォーマント（以下、外国人と記する）と日本人インフォーマント（以下、日本人と記する）の音声の「長さ」を分析し、比較してみた。

①外国人は「はちじゅう」の「じゅう」[jɯ:]を0.12秒の長さで発話しているが、日本人の[jɯ:]の長さは0.23秒である。藤崎・杉藤(1977)の日本人を対象にした知覚実験によると、短文における長母音と短母音の判断境界は0.17秒である。よって、この外国人の[jɯ:]は長母音と知覚されない可能性が高い。

②外国人の「まえに」の「に」[ni]の長さは0.33秒だが、日本人のそれは0.06秒で、外国人のほうが0.27秒長い。外国人のこの[ni]は長母音と知覚される可能性がある。

③外国人は、「おじいさん」の「お」[o]を言い淀みもあるため0.40秒と長く発話しており、長母音と知覚される可能性が高いが、日本人は0.09秒で、外国人のほうが0.31秒長い。

④外国人は「イタリア」の「リア」[ɾia]を0.08秒で発話しているが、日本人の[ɾia]は0.20秒で、外国人のほうが0.12秒短い。

⑤外国人は「わたし」の[t]の破裂前の持続時間を0.12秒とっている。これは、藤崎・杉藤(1977)の実験で明らかになった、促音の知覚判断境界0.16秒にはわずかにおよばないが、日本人の持続時間0.06秒より長く、ここに促音があると知覚される可能性がある。

⑥外国人の「ヨーロッパ」の[p]の破裂前持続時間は0.06秒である。これは、日本人の促音の知覚判断境界0.16秒に遠くおよばない長さであり、促音とは言えない。

⑦外国人は、「おなじ」の「な」[na]を0.19秒と長母音と知覚される可能性が高い長さで発話しているが、日本人の[na]は0.11秒であり、長母音と知覚される可能性は低い。

⑧外国人の「ほんとに」の「に」[ni]の長さは0.28秒であり、長母音と知覚される可能性が高いが、日本人の[ni]の長さは0.15秒であり、長母音と知覚されるかされないかの境界にある。外国人のほうが0.13秒長く発話している。

以上、述べてきたように、外国人インフォーマントの長さには、日本人がこの音声を聞いた際に長母音および促音の誤認識を引き起こす可能性がある音声上の問題が多く存在している。また、日本人インフォーマントの長さと大きく異なっていることも分かる。

なお、この外国人インフォーマントの音声には、高さおよび強さに関しても大きな問題が存在し、日本人と大きく異なっているが、本研究では、高さ・強さには手を加えずに合成を行うため、ここではその差異については詳述しない。

2－3－2．ポーズの分析・比較

次に、この二つの音声の「ポーズ」を分析・比較した。

①外国人は、「はちじゅうねん」と「まえ」の間に0.42秒のポーズを挿入しているが、日本人音声にはこの位置にポーズが存在していない。

②外国人は、「わたしの」から「おじいさん」の間に0.69秒の長いポーズを挿入しているが、日本人はポーズを挿入していない。
③外国人は、「いきましたね」と「きょう」の間のポーズを0.49秒とっているが、日本人のポーズは0.81秒と長い。日本人は文の間のポーズを十分にとっていることが分かる。
④外国人は「はんたい」の後にポーズを挿入せず、すぐに「わたしは」と続けている。一方、日本人は0.78秒の長いポーズを挿入している。
⑤外国人は「かえりますね」の後に全くポーズを挿入しておらず、文の間のポーズの重要性を認識していないようである。日本人は1.07秒のポーズを挿入しており、大きな差が見られる。
⑥外国人音声には「おなじもんだい」と「いつも」の間にポーズが存在していないが、日本人音声には0.36秒と短いがポーズが存在している。
⑦外国人は、「もんだいですね」と「おもしろい」の間にポーズを挿入していないが、日本人音声には0.92秒のポーズがある。
⑧外国人音声には、「ほんとに」と「たいへん」の間に1.39秒のポーズが存在するが、日本人のポーズは0.83秒と短く、0.56秒の差がみられる。

以上の分析の結果、外国人インフォーマントの音声には、ポーズに関しても長さ同様に大きな問題が存在しており、日本人インフォーマントの音声と大きく異なっていることが明らかになった。

2－4．合成音声の作成

次に、長さの影響力およびポーズの影響力の検証を行うため、外国人インフォーマントの実験音声をもとに合成音声を作成した。まず、長さの合成音声作成の手順は以下の通りである。

①外国人の発話文を印刷した用紙を日本人インフォーマントに読ませ、マイクを用いて音声分析器CSL4400に入力した。
②それぞれの音声をCSL4400の音声分析合成プログラム「シンセサイザ・プログラム5104」で分析し、数値データを得た。
③この数値データとWaveformおよび再生音声の聴解をもとにセグメント区分を行った。
④2名のセグメントの数値データを差し引きし、外国人の単音・ポーズ・高さ・強さは変えずに日本人の長さになるように数値データを入力し、合成した。

一方、ポーズの合成については、長さの合成の①②までは同様である。その後、

⑤Waveform上で、外国人・日本人双方のポーズ位置と長さを計測した。
⑥そのデータをもとに、外国人の音声のポーズを日本人と同様に変えた。具体的には、外国人・日本人双方において同位置にあるポーズにおいては、外国人のポーズ長を日本人のそれに変えた。また、外国人の音声中に日本人音声にないポーズがあれば消去し、逆

に日本人音声にあって外国人音声にないポーズがあれば、日本人の長さで挿入した。

以上の手順を経て、外国人インフォーマントの単音・ポーズ・高さ・強さは変えずに長さのみ合成した音声と、単音・長さ・高さ・強さは変えずにポーズのみ合成した音声の2つを得た。

2－5．評価表の作成

次に、実験の際に被験者に配布する評価表を作成した。評価表は、

質問①オリジナル音声は日本語音声として自然か

質問②長さを合成した音声は自然か

質問③ポーズを合成した音声は自然か

という3つの質問からなっている。質問①のオリジナル音声とは、何も手を加えていない外国人インフォーマントの音声のことである。そして、「自然か」という3つの質問は、7つの評価項目からなっている。まず、音声を聞いて「非常に不自然だ」と感じれば1、「どちらとも言えない」と感じれば4、そして「非常に自然だ」と感じれば7をチェックさせた。また、「不自然」から「どちらとも言えない」の中間であれば2または3、「どちらとも言えない」から「自然」の中間であれば5または6をチェックさせた。

以下に評価シートの1例を挙げる。

（例）　1　　2　　3　　4　　5　　6　　7

2－6．被験者および実験の手順

被験者は、大学生や日本語ボランティア、一般人など、計85名の日本人である。実験を行った時期は、2002年11月から2003年1月にかけてである。場所は、大学の講義室が多かったが、大学の教官研究室、居室など多岐にわたる。実験に要した時間は、事前説明も含め7,8分であった。

実験では、最初に評価表を配布した後、合成音声を用いた聴取実験であることを含め、入念な事前説明を行った。そして、質問①オリジナル音声→質問②長さを合成した音声→質問③ポーズを合成した音声　の順に聞かせて評価させた。

3．長さとポーズの合成音声の評価結果

3－1．各質問における項目間の差の検定

外国人インフォーマントのオリジナル音声と長さを合成した音声、およびポーズを合成した音声を日本人被験者がどのように評価したか、その結果について述べていく。

まず、3つの質問において、項目間に有意を持った差異が存在しているかを検証するため、SPSS for Windows 10.0.5Jを用いてχ 2乗検定を行った。その結果、質問①オリジ

ナル音声のχ2乗値は57.06、質問②長さの合成音声への評価のχ2乗値が16.58、質問③ポーズの合成音声への評価のχ2乗値が17.71であった。項目7(自由度6)の1%水準の臨界値が16.81、5%水準の臨界値が12.59であるため、質問①と③は1%水準 ($p \leqq .01$) で、質問②は5%水準($p \leqq .05$)で項目間に有意差があることが分かった。

3－2．評価結果

質問①．次の学習者の自然発話を聞いて、日本語音声として自然と感じれば7、非常に不自然と感じれば1、その中間と感じれば2～6をチェックしてください

この質問①に対する日本人被験者85名の評価結果を［表1］にあげる。

［表1］　オリジナル音声の自然さについての評価結果

	項目1	項目2	項目3	項目4	項目5	項目6	項目7
人数（人）	3	11	25	26	17	3	0
比率（%）	3.6	12.9	29.4	30.6	20.0	3.5	0

質問①に関しては、項目1から3を加えた「不自然度」が45.9%、項目5から7を加えた「自然度」が23.5%で、[何とも言えない]にあたる項目4が30.6%だった。評価は項目3・4・5に集中しているが、項目3が5より9.4ポイント(以下ptとする)多いこと、項目2が約13%いることから、全体的にみると「不自然度」が多いという評価になっている。

質問②．次の合成音声を聞いて、日本語音声として自然と感じれば7、非常に不自然と感じれば1、その中間と感じれば2～6をチェックしてください

［表2］長さの合成音声の自然さについての評価結果

	項目1	項目2	項目3	項目4	項目5	項目6	項目7
人数（人）	0	6	14	19	23	16	7
比率（%）	0	7.1	16.5	22.4	27.1	18.8	8.2

次に、［表2］の長さを合成した音声の評価結果だが、項目1から3を加えた「不自然度」が23.6%であり、項目5から7を加えた「自然度」の54.1%と比較すると30.5ptも少ない。詳細に見ていくと、項目1が0%、項目2も7.1%のみと少ない。逆に項目6が18.8%と質問①「オリジナル音声についての評価」と比して多く、項目7も8.2%いる。また、[何とも言えない]にあたる項目4は22.4%である。よって、オリジナル音声よりも「自然だ」という評価が多いと言える。

質問③．同じく、次の合成音声を聞いて、日本語音声として自然と感じれば7、非常に不自然と感じれば1、その中間と感じれば2～6をチェックしてください

［表３］ポーズの合成音声の自然さについての評価結果

	項目1	項目2	項目3	項目4	項目5	項目6	項目7
人数（人）	0	5	11	20	18	22	9
比率（%）	0	5.9	12.9	23.5	21.2	25.9	10.6

最後に［表３］のポーズの合成音声についてだが、こちらもオリジナル音声より「自然度」が高くなっている。項目5から7の「自然度」が57.7%で、「不自然度」の18.8%を38.9pt上回る。特に、[とても自然]にあたる項目７が10.6%あり、項目６が25.9%と全項目中最も多い。また、質問②同様、[とても不自然]の項目１は0%であり、[何とも言えない]の項目4は23.5%と質問②に近い。

３－３．それぞれの差の検定

次に、各質問の平均値の差に有意差が存在するかをみることにした。［表４］に、各質問における被験者の評価値（1から7）をもとにSPSSで算出した、外国人インフォーマントのオリジナル音声(質問①)と長さの合成音声(質問②)およびポーズの合成音声(質問③)の記述統計量をあげる。

［表４］質問①・②・③の記述統計量

	度数	最小値	最大値	平均値	標準偏差
質問①	85	1.00	7.00	3.6235	1.1850
質問②	85	2.00	7.00	4.6000	1.3645
質問③	85	2.00	7.00	4.8118	1.3844

まず、オリジナル音声と長さの合成音声、そしてオリジナル音声とポーズの合成音声の差をみるためにノンパラメトリック検定（Wilcoxonの符号付き順位検定）を行い、以下の結果を得た。［表５］に質問①(オリジナル)と②(長さ合成)の検定、［表６］に質問①と③(ポーズ合成)の検定結果をあげる。

［表５］質問①と②の検定統計量

質問②－質問①	
Z	－5.624（注）
漸近有意確率（両側）	0.0000000187

（注）負の順位にもとづく

［表６］質問①と③の検定統計量

質問③－質問①	
Z	－5.670（注）
漸近有意確率（両側）	0.0000000143

（注）負の順位にもとづく

このように、オリジナル音声と長さの合成音声、オリジナル音声とポーズの合成音声を比較したところ、ともに1%水準で有意差があることが分かった。［表４］にあるように、質問②・③ともに質問①を平均値で上回っているため、長さの合成音声・ポーズの合成音声ともに、有意差をもってオリジナル音声より自然だと評価されたことが分かる。

次に、長さの合成音声とポーズの合成音声の平均値の差に有意差があるかを、上記同様にノンパラメトリック検定を行って検証した。その結果を［表７］にあげる。

［表７］質問②と③の検定統計量

質問②－質問③	
Z	－2.068（注）
漸近有意確率（両側）	0.0386263877

（注）負の順位にもとづく

その結果、［表７］に見られるように、この２つの平均値の間には5%水準の有意差があることが明らかになった。［表４］にあるように、質問③ポーズの合成音声の平均値(4.8118)は、質問②長さの合成音声の平均値(4.6000)を上回っているので、今回の被験者は、ポーズを合成した音声のほうが、長さを合成した音声より有意差をもって自然だと評価したと言える。つまり、単音・高さ・強さを変えずに合成して作成された２つの音声の評価において、ポーズを合成した音声のほうが長さを合成した音声よりも、自然かどうかという評価に与える影響力が大きいということである。

３－４．まとめ

今回は、長さとポーズが評価に与える影響力を、聴取実験を通して検証した。そして、ポーズを合成した音声の評価が、長さを合成した音声の評価を上回るという結果を得た。これは、ポーズのほうが長さよりも自然かどうかという評価に与える影響力が大きいということを意味する。今回の結果から「ポーズは長さよりも重要だ」と一般化することはあまりに早急に過ぎるが、これまであまり注目されてこなかった「ポーズが音声評価に与える影響力」について再考させられる結果だと言える。

４．ポーズ指導について

上述してきたように、ポーズが影響力の大きい音声項目であることが明らかになった以上、今後は、高さや長さ同様、ポーズをも重視して注意しつつ教える必要がある。ポーズは、単語レベルではさほど重要な関わりを持たないが、文レベルにおいてその重要性を増してくる。特に、長い文章レベルでは、適切なポーズが挿入されているかいないかは聞き手の認識しやすさに大きな影響を与える。よって、これまでも頻繁に行われてきた単語レベルでの母音の長短やアクセントの指導だけでなく、文レベル・文章レベルでのポーズ指導の徹底が望まれよう。特に、初級レベルの学習者に対しては不要なポーズの挿入回避を意識させ、そのことが「スムーズさ」につながることを指導するべきである。また、中上級レベルの学習者に対しては、文節間または文の間の適切なポーズの挿入を指導し、ポーズ挿入による聞き手側の認識しやすさの向上を情報として伝えるべきである。

さらに、他の音声項目同様、自らのポーズに問題があるかどうかは、学習者が発話しながらでは認識しにくい。そこで、学習者に自らのポーズの問題を認識させるためには、発話を録音した後に学習者に聞かせ、ポーズの問題を自らモニターさせる指導法が有効であ

ると考える。この自己モニターを繰り返すことにより、学習者は「スムーズさ」につながるポーズの重要性を認識するようになり、自然発話をする際にもポーズに注意するようになる。この注意が無意識に行われるようになれば、適切なポーズ挿入を含む自然な日本語音声習得に近づいていくのではないだろうか。

参考文献

皆川泰代・前川喜久雄・桐谷滋　2002　「日本語学習者の長/短母音の同定におけるピッチ型と音節位置の効果」
『音声研究』第6巻第2号　pp.88-97

片桐　恭弘　2000　「音声メタコミュニケーション」
『音声研究』第4巻第3号　pp.57-59

中川道子・二村年哉　2000　「初級日本語学習者の長短母音の認識傾向と持続時間」
『北海道大学留学生センター紀要』第4号　pp.18-39

原田　明子　1998　「一般の日本人は外国人の日本語をどのように評価するか」
『北海道大学留学生センター紀要』第2号　pp.157-168

杉藤美代子　1997　「話し言葉のアクセント・イントネーション・リズムとポーズ」
『アクセント・イントネーション・リズムとポーズ』三省堂　pp.3-20

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

水谷　　修　1993　「ポーズ」
『日本語教育ハンドブック』大修館書店　pp244-245

藤崎博也・杉藤美代子 1977 「音声の物理的性質」
『岩波　講座日本語 5　音韻』岩波書店　pp63-108