

Title	聴知覚メカニズムに着目した音声の年齢知覚に関する研究
Author(s)	畠山, 達也
Citation	
Issue Date	2019-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/15908
Rights	
Description	Supervisor: 鷗木 祐史, 先端科学技術研究科, 修士 (情報科学)

修 士 論 文

**聴知覚メカニズムに着目した
音声の年齢知覚に関する研究**

1710163 畠山 達也

主指導教員	鵜木祐史
審査委員主査	鵜木祐史
審査委員	赤木正人
	党建武
	吉高淳夫

北陸先端科学技術大学院大学
先端科学技術研究科 [情報科学]

平成 31 年 2 月

Abstract

Many people can correctly estimate the age of a speaker simply by listening to the speaker's voice. Age perception can be regarded as the perception of nonlinguistic information, so it has been studied extensively of speaker identification and speaker individuality. If we can reveal the mechanism behind auditory perception in terms of how individuals perceive age from an observed voice, this finding will contribute to research on speaker identification, and also to the applications of various speech signal processing techniques such as age estimation. The main problem in studying speech science related to age perception is clarifying the relationship between the corresponding acoustic features on the basis of the speech production mechanism and age perception from a speaker's voice. In order to reveal the nature of age perception, it is necessary to investigate it from the viewpoint of speech production. In this paper investigate the relationship between the perceived age of a speaker's voice and the corresponding auditory factors (roughness, fluctuation strength, and sharpness) sound quality metrics.

I conducted experiments in which i labeled the perceptual ages to corresponding speech stimuli in three data based. The first speech corpus, CIAIR-VCV, includes the speech of men and women aged 6 to 12. The speech signals of all 14 speakers were used as stimuli in the child speech dataset. The second corpus, APP6BLA, includes thirty male and female speech of teens to 60s. The speech signals of 74 speakers were used as stimuli in the adult speech dataset. The third corpus, S-JNAS, includes male female voice over 60's. The speech signals of 28 speakers were used as stimuli in the adult speech dataset. Age perception experiments was carried separately for each dataset. Ten native japanese speakers with normal hearing (six males in their 20s, four female in her 20s) participated in the experiment. The both experiments were carried out in a soundproof room. Speech stimulus was randomly presented. The participants were required to answer the perceived age using values of one year. As the results, it was shown that the listener accurately estimates the age of the speaker to some extent. Whether it is a male speaker or a female speaker, the voice of the child is smaller or smaller than the adult. In the case of adult speech, the variation of the result becomes large, and the perceived age is estimated to be lower than the actual age. In the case of women, this tendency appeared more prominently.

Previous studies have shown that the spectral tilt is highly correlated with actual age in the case of male speech, and the fundamental frequency is highly correlated with actual age in the case of female speech. So, i investigated the relationship between these two acoustical features of the speech stimuli that we used in the experiments and perceptual age. The spectral tilt in the dB per octave frequency was determined using the least

mean squared method (LMS) from shortterm. Here, fourier transformation in which the window function is Hanning, frame length is 10 ms, and frame shift is 5 ms. The F0 was obtained using STRAIGHT TEMPO. F0 and spectral tilt were then averaged in the time domain to use them as the acoustical features.

From the result, looking at the ages as a whole, in the case of male voice, there is a tendency that roughness increases as the perceived age increases. On the other hand, looking at children, adults, and elderly people, roughness tends to decline in elderly counties. In the case of female voice, there was no tendency in roughness to perceived age when looking at the ages as a whole. However, when divided into small age groups, descending tendency was seen in child counties and elderly counties. Relationship between sensory age and fluctuation strength of male speaker, relationship between perception age and fluctuation strength of female speaker. Looking at the ages as a whole, in the case of male voice, there is a tendency that the perceived age increases and the fluctuation strength decreases. On the other hand, when divided into children, adults, and elderly people, the tendency of roughness to decline was observed in children and adults, and the trend of fluctuation strength tended to increase in elderly counties. In the case of female speech, there was no tendency in the fluctuation strength with respect to perceived age as seen in the entire age. However, when divided into small age groups, a downward tendency was seen in child counties. Relationship between perceived age of male speaker and sharpness, from the result of the relationship between perception age and sharpness of female speaker In the case of male speech, there was no tendency in sharpness to perceived age. When dividing into children, adults, and elderly people, Sharpness was seen to rise in elderly people. There was no tendency in sharpness against sensory age in female speech. However, descent tendency was seen in child counties.

From the results, in the case of male voice, there is a tendency that the perception age increases and the roughness increases. On the other hand, when divided into children, adults, and elderly people, roughness tends to decline in elderly counties. In the case of female speech, there was no tendency to roughness to perceived age. However, descent tendency was seen in child counties and elderly counties. Relationship between perceived age and fluctuation strength of male speaker, from the results of the relationship between perceived age and fluctuation strength of female speaker. In the case of male voice, there is a tendency that the perceived age increases and the fluctuation strength decreases. On the other hand, when divided into children, adults, and elderly people, the tendency of roughness to decline was observed in children and adults, and the trend of fluctuating intensity was seen in the elderly counties. In the case of female speech, there was no tendency in the fluctuation strength with respect to perceived age. However, descent tendency was seen in child counties. Relationship between perceived age of male speaker and sharpness, from the result of the relation between perceived age of female speaker

and sharpness. In the case of male speech, there was no tendency in sharpness against perceived age. On the other hand, when divided into children, adults, and elderly people, sharpness was seen to rise in elderly people. In the case of female speech, there was no tendency in sharpness to perceived age. However, descent tendency was seen in child counties.

Comparing the case of seeing the relation between perceived age and acoustic feature and the case of seeing the relationship between perceived age and sound quality metrics, In men, F0 ($r^2 = 0.57$) was the most highly correlated with the overall perceived age of children, adults, and the elderly. The next highest was the spectral tilt ($r^2 = 0.42$). In the case of women, it was F0 ($r^2 = 0.23$). The next highest was spectral tilt and sharpness ($r^2 = 0.21$). However, since the spectral tilt varies widely, other feature quantities may be involved.

Based on the perceptual age and spectral tilt, and the relationship between perceived age and F0, as the tendency was seen in perceived age and spectral tilt of male voice, it is possible to explain the result of this time with the acoustic feature which is conventionally known for perceived age. However, the value of correlation to perceived age is low. Therefore, regarding perceptual age, it is necessary to evaluate the impression of hearing. From the result of the relation between perceived age and sound quality metrics, i could not find an indicator that could explain the whole age. Therefore, further analysis was attempted by considering fluctuation of perceived age. From the results in Chapter 4, i extracted 10 data with small standard deviations and obtained relationships with each feature quantity. As a result, the perceptual age and the roughness could be related.

目次

第1章 序論	1
1.1 はじめに	1
1.2 研究背景	1
1.3 研究の目的	2
1.4 本論文の構成	2
第2章 年齢に関わる研究	5
2.1 加齢に伴う音響特徴	5
2.2 音声生成のエイジング	5
2.3 年齢知覚	6
第3章 本研究の方略	8
3.1 本研究の着想点	8
3.2 音質評価指標	8
3.3 研究の方法論	9
第4章 年齢知覚実験	11
4.1 本実験の目的	11
4.2 実験用データベース	11
4.3 実験方法	15
4.4 実験結果	16
4.5 考察	19
第5章 年齢知覚に関わる特徴の分析	20
5.1 分析方法	20
5.2 知覚年齢と音響特徴の関係	20
5.3 知覚年齢と音質評価指標の関係	26
5.4 音響特徴と音質評価指標の結果の比較	33
5.5 考察	33
第6章 全体考察	34

第7章 結論	35
7.1 本研究で明らかになったこと	35
7.2 残された課題	35
参考文献	36
研究業績一覧	39
謝辞	40

目 次

1.1	論文の構成	4
4.1	男性話者の各年齢に対する出現頻度	12
4.2	女性話者の各年齢に対する出現頻度	13
4.3	実験環境	15
4.4	実年齢と知覚年齢の関係（男声）	17
4.5	実年齢と知覚年齢の関係（女声）	18
5.1	知覚年齢とスペクトル傾斜の関係（男声）	22
5.2	知覚年齢とスペクトル傾斜の関係（女声）	23
5.3	知覚年齢と F0 の関係（男声）	24
5.4	知覚年齢と F0 の関係（女声）	25
5.5	知覚年齢とラフネスの関係（男声）	27
5.6	知覚年齢とラフネスの関係（女声）	28
5.7	知覚年齢と変動強度の関係（男声）	29
5.8	知覚年齢と変動強度の関係（女声）	30
5.9	知覚年齢とシャープネスの関係（男声）	31
5.10	知覚年齢とシャープネスの関係（女声）	32

表 目 次

4.1 発話内容	14
--------------------	----

第1章 序論

1.1 はじめに

人は、話者の声を聴いただけでも話者の年齢をある程度正確に推定することができるだろう。人の声には性別や年齢などといった話者の特徴である個人性が含まれている。人はこの個人性を用いて話者の弁別を行っている。そして、感じた相手の特徴をもとに自分自身の話し方等を決めている。これは日常的に行っており、例えば電話口の相手が何歳か考え話し方を変えたり、相手が誰なのかを考えたりするだろう。このときもやはりある程度は正確に年齢あてができていよう。しかし、人はどのようなメカニズムで年齢を知覚しているのだろうか。

これまで加齢と声帯や声道の構造の変化の関係を調べられてきた。しかし、年齢知覚のメカニズムは未だ明らかになっていない。声の年齢知覚は、音声の個人性知覚の一つとして深く検討されている。そのため、音声科学の研究として声の年齢知覚を検討することは、個人性の知覚メカニズムの解明にも役立つはずである。また、年齢知覚のメカニズムを明らかにすることが出来れば、機械による年齢推定や個人性、声質判断の実現にも貢献できるものと考えられる。

年齢知覚に係る音声科学の研究では、音声生成のメカニズムに基づき、音声生成系の加齢による構造変化から声の年齢知覚とそれに係る音響特徴を解明することが主要な課題である。例えば、従来研究の検討結果から、音声生成系の変化に係る音響特徴の変化を検討することで話者の年齢知覚に関する検討が行われてきた。その結果、加齢に伴い、基本周波数 F_0 やフォルマント周波数、スペクトル傾斜、話速といった音響特徴の変化が重要であることが分かった。また、声の知覚年齢が話者の実年齢と異なり、聴取者ごとに個人差があることも分かった。

1.2 研究背景

声の年齢知覚に関して「ことばの鎖」を抛りどころに、加齢による聴力低下と音声変形の相互作用として高齢者の実年齢と音声刺激に対する聴取者によって知覚された年齢(知覚年齢)の関係が水町によって調べられてきた。その結果、声の実年齢と知覚年齢の差の関係が説明されているが、年齢知覚に係る聴覚的要因の解明には至っていない。

一方、咽喉を専門とする医師は、話者の声から病理音声を容易に判断することができる。特に、病理音声の判別には、嗄声度を点数化する指標として、GRBAS 尺度が用いら

れている。特定の用語を使わずに声質の非類似性を評価した場合、聴取者は程度の差はあるものの GRBAS 尺度で用いられている嗄声の重症度を表す Grade や粗造性・ガラガラ感を表す Roughness、気息性・カサカサ感・ハスキーボイスを表す Breathy、無気力性・弱々しさを表す Asthenic、努力性・力みを表す Strained といった概念を利用していることが今泉によって報告されている。また、健常者、喉頭癌患者、反回神経麻痺患者、声帯ポリープ患者の音声を利用して、変動周期指数 (PPQ) や規格化雑音エネルギー (NNE) と GRBAS 尺度との関係を調査した報告がある [5]。その結果、PPQ や NNE を対数表現したものと GRBAS 尺度の内のガラガラ感に対応する Roughness との相関が高いことも報告されている。

1.3 研究の目的

本研究の目的は、聴覚的要因として音色知覚に係る音質評価指標（ラフネス、変動強度、シャープネス）を用いて、人の年齢知覚に関する知見を深めることである。ここで、ラフネスは音の粗さを、変動強度は音の変動感の強さを、シャープネスは音の鋭さ・甲高さを表す感覚尺度であり、聴覚系の周波数選択性・変調知覚特性から得られる音響特徴に深く関係する。本研究では、子供から高齢者までの幅広い年齢層の音声データに対し、各指標の影響度を調査することで、人がなにを年齢知覚の手がかりにしているか調査する。

1.4 本論文の構成

本論文は、7章で構成される。図 1.1 に本論文の構成を図示する。

第1章

この章は序論であり、本論文で対象とする年齢知覚に関する研究背景述べる。これにより、研究の目的を明らかにする。

第2章

この章では、年齢に関わる研究を紹介する。まず、加齢に伴う音響特徴の変化について明らかにされた先行研究を紹介し、次に、音声生成系のエイジングに関わる研究を紹介する。最後に年齢の知覚に関わる研究を紹介する。

第3章

この章では、研究のアプローチを述べる。本研究の着想点を説明した上で、本研究の新規性でもある音色知覚に関わる音質評価指標についての説明を行う。その後、本研究の方法論を説明する。

第4章

この章では、知覚年齢のラベル付のために行った聴取実験について述べる。実験では子供から高齢者までの文章発話音源を利用して実年齢に対する知覚年齢を求める。それによって、実年齢と知覚年齢の関係を検討する。

第5章

この章では、知覚年齢と従来研究で実年齢との関係が示唆された音響特徴 (F_0 とスペクトル傾斜) との関係について求めた分析の結果を示した上で、知覚年齢と本研究の特色である音質評価指標との関係について求めた分析の結果を示す。その後、各分析の結果を比較する。

第6章

この章では、実年齢と知覚年齢の関係、知覚年齢と音響特徴の関係、知覚年齢と音質評価指標の関係を関連付けて全体的な考察を述べる。

第7章

この章は結論であり、本研究で明らかになったことを明らかにした上で、今後の展望について述べる。

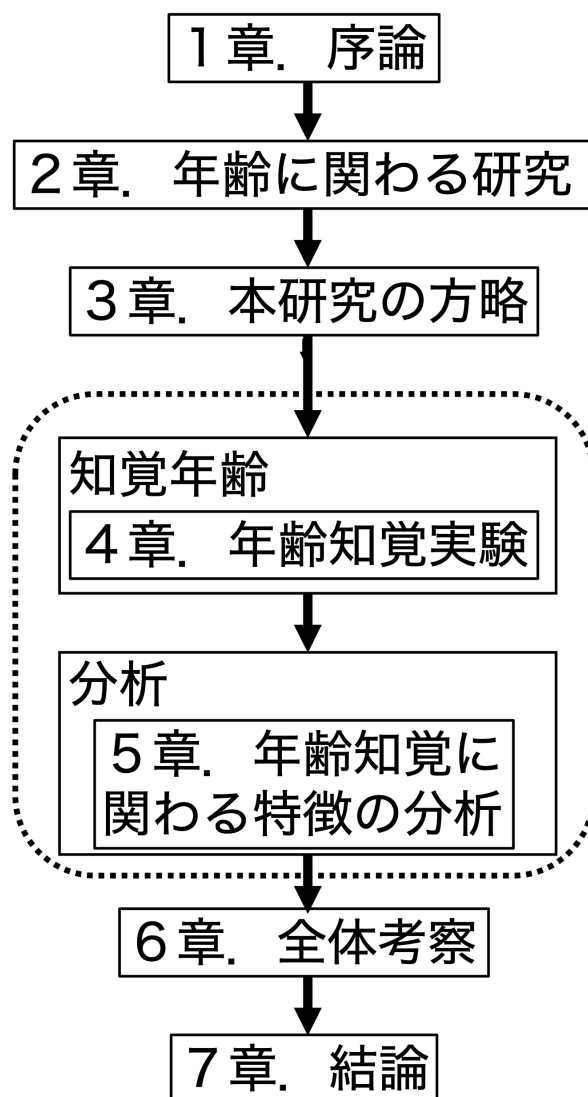


図 1.1: 論文の構成

第2章 年齢に関わる研究

2.1 加齢に伴う音響特徴

人の声は加齢により変化する。従来研究では、音声生成に係る音響特徴の変化が検討されてきた。特に男性では話速や高周波帯域のスペクトルが、女性では基本周波数が変化することが報告されている [1] [2]。また、男児は成長過程で喉頭下降などにより基本周波数が急激に下降することが報告されている [1]。森らは、6歳から20歳までの男女の音声の実年齢とF0やホルマント周波数の関係を調べた [2]。男性のF0は年齢とともに降下することがわかった。変声期を迎えたあとで急激に降下し、60代を過ぎると上昇する傾向が得られた。女声の場合は急激な変化ではなく、徐々に降下し成人のF0に近づく傾向が得られている。加えて、粕屋らは、男性話者29名、女性話者59名のF0について10年から最大24年間にわたって追跡した結果を報告している [3]。その結果、女性話者の場合、加齢に伴い下降する傾向が得られている。また、加齢に伴い基本周期のゆらぎ (APQ) が増加する傾向も得られた。そして、男女ともに1周期内の最大振幅の周期ごとのゆらぎ (PPQ) が増加する話者が多いと報告している。また、女性話者では全周波数帯域の喉頭雑音量 (NNE) が増加することも報告されている。青年群 (19～29歳) の音響特徴と老年群 (60～78歳) の音響特徴を比較した研究もある [4]。この研究では、男性話者の場合、粕屋らの研究と同様にF0が上昇し、PPQも上昇する傾向が得られている。この研究では、PPQなども上昇する傾向が得られている。女性の場合は、F0は減少する傾向が得られている。

これらの結果からも加齢とF0の関係性が示唆されている。

2.2 音声生成のエイジング

音声は、声帯振動によって生じた波が声道を通過し、空気中に放出される事によって生じる。そのため、加齢による音声生成系のエイジングが発話者の声の特徴に影響を及ぼす。ここでは、音声生成系のエイジングに関する先行研究と音声生成系のエイジングに少々関わりがあると考えられる病理音声に関する先行研究について紹介する。まず、加齢と声帯の関係を調べた研究がある。新生児から69歳までの男女の声帯を調べた研究によると声帯並びに声帯膜様部は新生児から加齢とともに増大していき、20歳頃に成人の長さに達するとされている [5] [6]。声帯の長さは、9歳頃までは男女に差はあまりないが、15歳以降では男性のほうが長くなるということも分かった。また、70～104歳の男女の声帯を調べた追研究によると、声帯粘膜全体の厚さに関して、女性では70歳以降で粘膜が暑くな

る傾向が示され、男性は一定の傾向は得られないと示された [7]。そして、声帯膜様部に対する粘膜の厚さの比は、加齢とともにどう題する傾向があり、特に女性において顕著であるという結果が得られている。次に、加齢と声道形状の関係を調べた研究がある [8]。子供の声道発達について調べた研究によると、子供の声道の大きな特徴として、大人に比べ短いことが示されている [9]。生まれたときの声道長が 8 cm なのに対して最終的には 17 cm ほどに伸長することも示されている。また、2～25 歳の年齢に伴う声道長の変化を観測した研究がある。その結果、思春期前から思春期にかけて声道のすべての部位に関して伸長を確認したが、思春期後では軟口蓋と喉頭に関してのみの変化であると明らかにされた [9]。

音声生成系のエイジングに関わるもので、病理音声というものが、咽喉を専門とする医師は、話者の声から病理音声を容易に判断することができる。特に、病理音声の判別には、嗄声度を点数化する指標として、GRBAS 尺度が用いられている。特定の用語を使わずに声質の非類似性を評定した場合、聴取者は程度の差はあるものの GRBAS 尺度で用いられている嗄声の重症度を表す Grade や粗造性・ガラガラ感を表す Roughness、気息性・カサカサ感・ハスキーボイスを表す Breathy、無気力性・弱々しさを表す Asthenic、努力性・力みを表す Strained といった概念を利用していることが今泉によって報告されている [10]。また、健常者、喉頭癌患者、反回神経麻痺患者、声帯ポリープ患者の音声を利用して、PPQ や NNE と GRBAS 尺度との関係を調査した報告がある [11] [12]。その結果、PPQ や NNE を対数表現したものと GRBAS 尺度の内のガラガラ感に対応する Roughness との相関が高いことも報告されている。先行研究によって、音声生成系は加齢によって声帯も声道も変化することが明らかにされた。これらの特徴を人は聞いて年齢知覚の手がかりにしている可能性はある。また、病理音声で用いられるような特徴も年齢の知覚に関係があるのかもしれない。

2.3 年齢知覚

年齢の知覚に関して調べられた研究は多い。声の年齢知覚に関して「ことばの鎖」を抛りどころに、加齢による聴力低下と音声変形の相互作用として高齢者の実年齢と音声刺激に対する聴取者によって知覚された年齢 (知覚年齢) の関係が水町によって調べられてきた [13]。その結果、高齢者の音声では、4.5 kHz 以上の高周波帯域でのパワーが上昇していることが確認された。また、パワースペクトルのダイナミックレンジが縮小していることも確認された。そして、加齢による純音聴力閾値の上昇は 4 kHz 及び 8 kHz の上昇が特徴的であることが示された。これらの結果には相補関係があると説明されている。また、高齢者音声の年齢の知覚を調べたところ、難聴者の音声は、実年齢よりも老けて知覚される傾向があり、正常聴力を有する話者の音声は、実年齢よりも若く知覚される傾向が示された。また、知覚年齢が実年齢よりも老けて知覚された高齢者の音声の印象は、嗄声により支配的であるということも明らかにされた。また、知覚年齢と基本周波数の関係を調査した研究がある [14]。その結果、20 代の女性という狭い範囲ではあるが知覚年齢に影響

があることが示唆された。これらの他に，感情音声や歌声，音声発生形態に関わる年齢知覚の研究なども盛んに行われた [15] [16] [17] [18]。

一方，音声認識技術を用いて知覚年齢の推定を行った研究もある。峯松らは，GMM 話者モデリング技術を利用して，知覚年齢の推定を試みた [19] [20]。その結果，実際に人が推定した知覚年齢と GMM 話者モデリング技術を用いて推定した知覚年齢に高い相関が得られたと報告している。また，主観的高齢者の同定では，91 %の正答率が得られる結果となった [21] [22]。

第3章 本研究の方略

3.1 本研究の着想点

年齢知覚に関しては、これまで音声生成器官の加齢による変化や加齢に関わる音響特徴の変化から調べられてきた。しかし、年齢知覚の本質を知るためには、音声生成だけではなく聴知覚からの検討も必要である。ガラガラ感や変動感が病理音声の判断に深く関係することを考えると、これらの指標を加齢による声質の変化、特に年齢を測ることに使うことができるかもしれない。さらには、ラフネスや変動強度、シャープネスといった音質評価指標を利用して知覚年齢を推定できるかもしれない。

3.2 音質評価指標

車や工場機械などの騒音計測、異音検査などに用いられてきた指標に音質評価指標というものがある。音質評価指標は指標ではあるが、聴覚フィルタバンクをイメージして、聴覚フィルタの臨界帯域でのフィルタ数が35ポイントとして、1 BARK から 35 BARK までの帯域のラウドネス密度を見ている。その値を使ってラフネスや変動強度、シャープネスが定義されている。例えば、聴覚的なスペクトルを求めて高域に重み付けして出できた重心がシャープネス、この研究では定常だと考えて時間平均。ラフネスは聴覚フィルタバンクから出てくる時間方向の変動を見たときに、それがどのくらいばらついているかをしらべた統計量になっていて、それに各チャンネルの平均値として出てくるものがラフネス。その周波数の上限下限によってラフネスと変動強度に分けられる。やはり、いずれも聴覚フィルタバンクを利用している。聴覚フィルタをイメージしていて音色知覚を考えたときに聴覚フィルタバンクを考えて、出てきた出力の中から以下のような計算をしたものが音質評価指標である。その出力から音色に関わる特徴をみるものとしてラフネスや変動強度、シャープネスがある。このように聴知覚のメカニズムに沿って算出された値としてラフネスや変動強度、シャープネスを利用する。

まず、ラフネスは音のざらつき感を表す指標である。ラフネスは、音圧レベルが60 dB、搬送波周波数が1 kHz、変調周波数が70 Hz、変調度が1のとき、1 asper と定義されており、次式のように表される [23] [24]。

$$R = 0.3f_{\text{mod}} \int_0^{\text{Bark}} \Delta L(z) dz, \quad (3.1)$$

ただし、 f_{mod} は変調周波数、 ΔL は臨界帯域ごとの興奮レベルの山と谷の差、 z は臨界帯域番号である。変動強度は音の変動感を表す指標である。変動強度は音圧レベルが 60 dB、搬送波周波数が 1 kHz、変調周波数が 4 Hz、変調度が 1 のとき、1 vacil と定義されており、次式のように表される [23] [24] [25]。

$$F = \frac{0.008 \int_0^{\text{Bark}} \Delta L(z) dz}{\frac{f_{\text{mod}}}{4} + \frac{4}{f_{\text{mod}}}}, \quad (3.2)$$

ただし、 f_{mod} は変調周波数、 ΔL は臨界帯域ごとの興奮レベルの山と谷の差、 z は臨界帯域番号である。シャープネスは音の甲高さを表す指標である。シャープネスは、音圧レベルが 60 dB で、1 kHz を中心周波数とする臨界帯域幅の狭帯域雑音のときに 1 acum と定義されており次式のように表される [23] [24]。

$$S = 0.11 \frac{\int_0^{\text{Bark}} N'(z) g'(z) dz}{\int_0^{\text{Bark}} N'(z) dz}, \quad (3.3)$$

ただし、 $N'(z)$ は臨界帯域ごとのラウドネス（ラウドネス密度）、 $g'(z)$ はシャープネスの重み、 z は臨界帯域番号である。

3.3 研究の方法論

本研究では、4 章において、発話者の実年齢と聴取者が知覚する知覚年齢の関係を明らかにするために、子供、大人、高齢者の 3 種類のデータベースを統合したうえで実年齢に対する知覚年齢のラベル付を行う聴取実験を行う。聴取実験では発話内容を年齢推定の手がかりにしないために発話内容を考慮したうえで、実験、分析に用いる音声刺激の選定を行う。実年齢と知覚年齢の関係を明らかにすることができれば、個人性の知覚メカニズムの解明を深めることにもつながると考えられる。そして、5 章においては、4 章で得られた知覚年齢を用いて、まず、先行研究で実年齢との関係が明らかにされた音響特徴と知覚年齢との関係を回帰分析を用いて確認する。2 章で説明したように、加齢に伴って様々な音響特徴が変化する。本研究では、加齢に関わる音響特徴の内 2 種の音響特徴を選定した。女性の実年齢との関係が特に示唆された F0 と男性の実年齢との関係が示唆され、聴力低下と音声変化の相互作用としても影響があったと考えられるスペクトル傾斜を利用することとした。次に、本研究の特色でもある聴覚的要因として音色知覚に係る音質評価指標（ラフネス、変動強度、シャープネス）を用いて知覚年齢との関係を見ることで、人がどういう特徴を年齢知覚の手がかりにしているかを明らかにしていく。これによって年齢知覚に関する知見を深めること延いては個人性知覚メカニズムの解明につながると考えられる。最後に、6 章では 4 章の実年齢と知覚年齢の関係を示す結果と 5 章の従来研究で加齢との関係を示唆された音響特徴や本研究の新規性、重要性にもなる聴覚的要因として

音色知覚に係る音質評価指標（ラフネス，変動強度，シャープネス）と知覚年齢の関係性の結果を合わせて総合的に考察をすることで音声の年齢知覚に関する知見を深めていく．

第4章 年齢知覚実験

4.1 本実験の目的

本章では，本研究で利用する音声刺激に対する発話者の実年齢と知覚年齢の関係を明らかにするために，音声刺激に対する知覚年齢のラベル付けの聴取実験を行う．

4.2 実験用データベース

実験には，子供の声データベース（CIAIR-VCV）および ATR 多数話者データベース（APP6BLA），新聞記事読み上げ高齢者音声コーパス（S-JNAS）に収録された音声データを利用した [26] [27] [28]．0 歳から 79 歳まで各実年齢ごとに男女 1 名ずつのデータをランダムに選択した．尚各年齢に対応する音声刺激がない場合もあった．図 4.1 には，男性話者の各年齢に対する音声刺激の出現頻度を，図 4.2 には，女性話者の各年齢に対する音声刺激の出現頻度を示す．CIAIR-VCV には 6 歳から 12 歳の男女の音声 that 収録されており，分析では男性話者 7 名，女性話者 7 名の計 14 名の音声を実験刺激として利用した．また，APP6BLA には 10 代から 60 代の男女の音声 that 収録されており，分析では男性話者 30 名，女性話者 44 名の計 74 名の音声を実験刺激として利用した．そして，S-JNAS には 60 代以上の男女の音声 that 収録されており，分析では男性話者 14 名，女声話者 14 名の計 28 名の音声を実験刺激として利用した．本研究で利用した音声刺激の発話内容を表 4.1 に示す．表 4.1 に示すように大人と高齢者のデータベースは発話内容が同じものを選択できている．しかし，子供のデータベースにおいては，同様の発話をしているものがなかったために発話時間を考慮して童話文章発話音声を採択した．

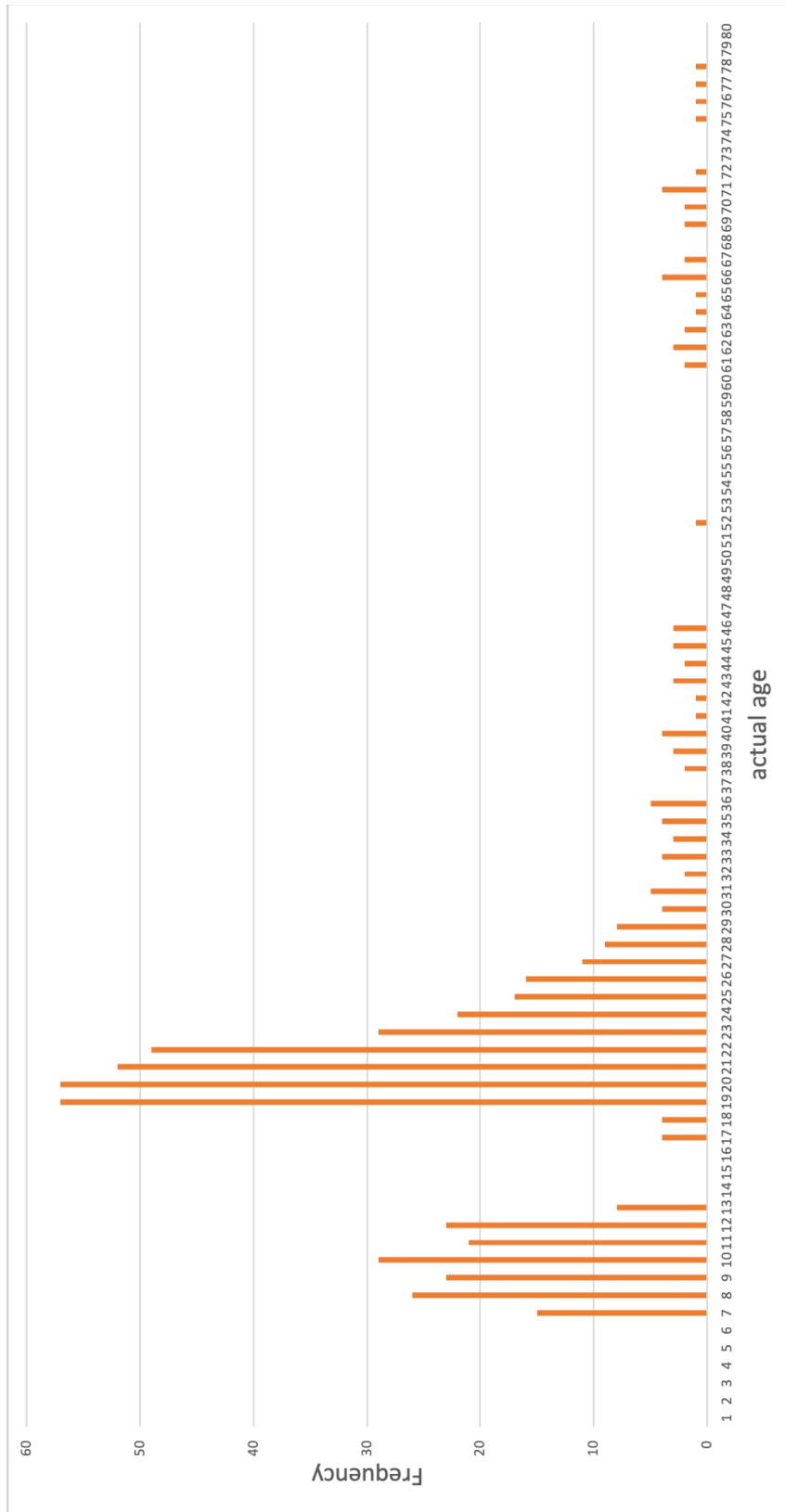


図 4.1: 男性話者の各年齢に対する出現頻度

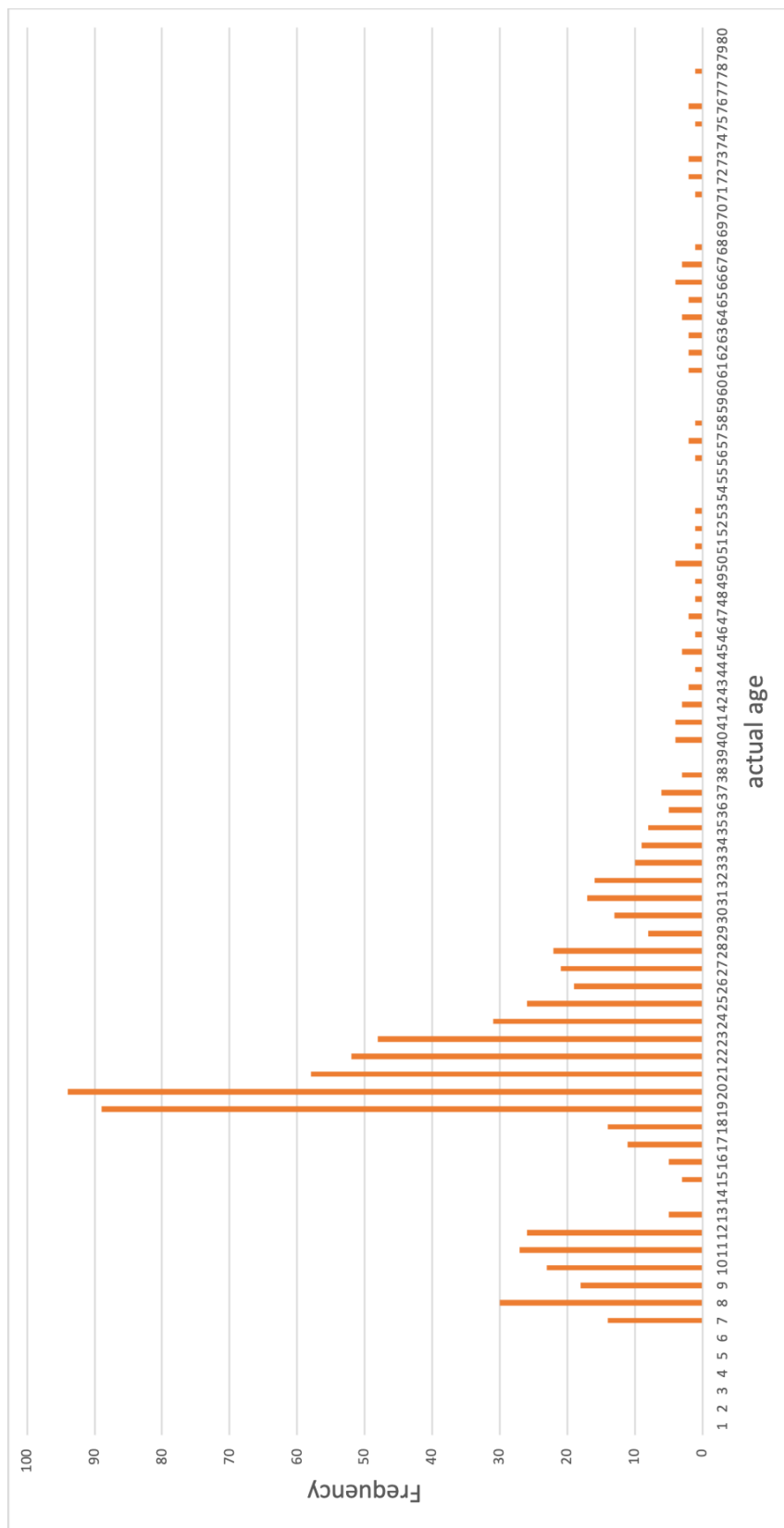


図 4.2: 女性話者の各年齢に対する出現頻度

表 4.1: 発話内容

	子供	大人・高齢者
1	あたりはもうまっくらで、 こんこんと雪が降っていました。	一週間ばかりニューヨーク取材した。
2	少女はもっとあたたまろうと、 足をのばしました。	会場の周辺には原宿駅や代々木駅もあるし、 ちょっと歩けば新宿御苑駅もある。
3	少女はもう一度マッチをすりました。	老人ホームの場合は、健康器具やひざ掛けだ。

4.3 実験方法

実験では正常聴力を有する日本語母語話者 10 名（20 代男性 6 名，20 代女性 4 名）が参加した。実験は防音室内で行われた。実験参加者には，ヘッドホンにて男性音声の場合は 153 音声をランダム呈示し，女性音声の場合は 186 音声をランダムに提示し，知覚した年齢を 1 歳刻みで回答させた。回答は GUI 上にキーボードから直接入力させた。ヘッドホンからの出力レベルは，B&K NEXUS, B&K type 2231 モジュール型精密騒音計を利用して，実験前に毎回聞きやすいレベル（約 65 dB 前後）に校正された。本実験では，年齢知覚をする際にまず男女どちらの音声であるかを考えるのではないかと考え，実験前に発話者が男性であるかもしくは女性であるかを教示した。また実験用データベースの説明に記したように子供の音声と大人，高齢者の音声とでは発話内容が異なるため，実験前に発話内容を年齢知覚の手がかりにしないように事前注意をした。

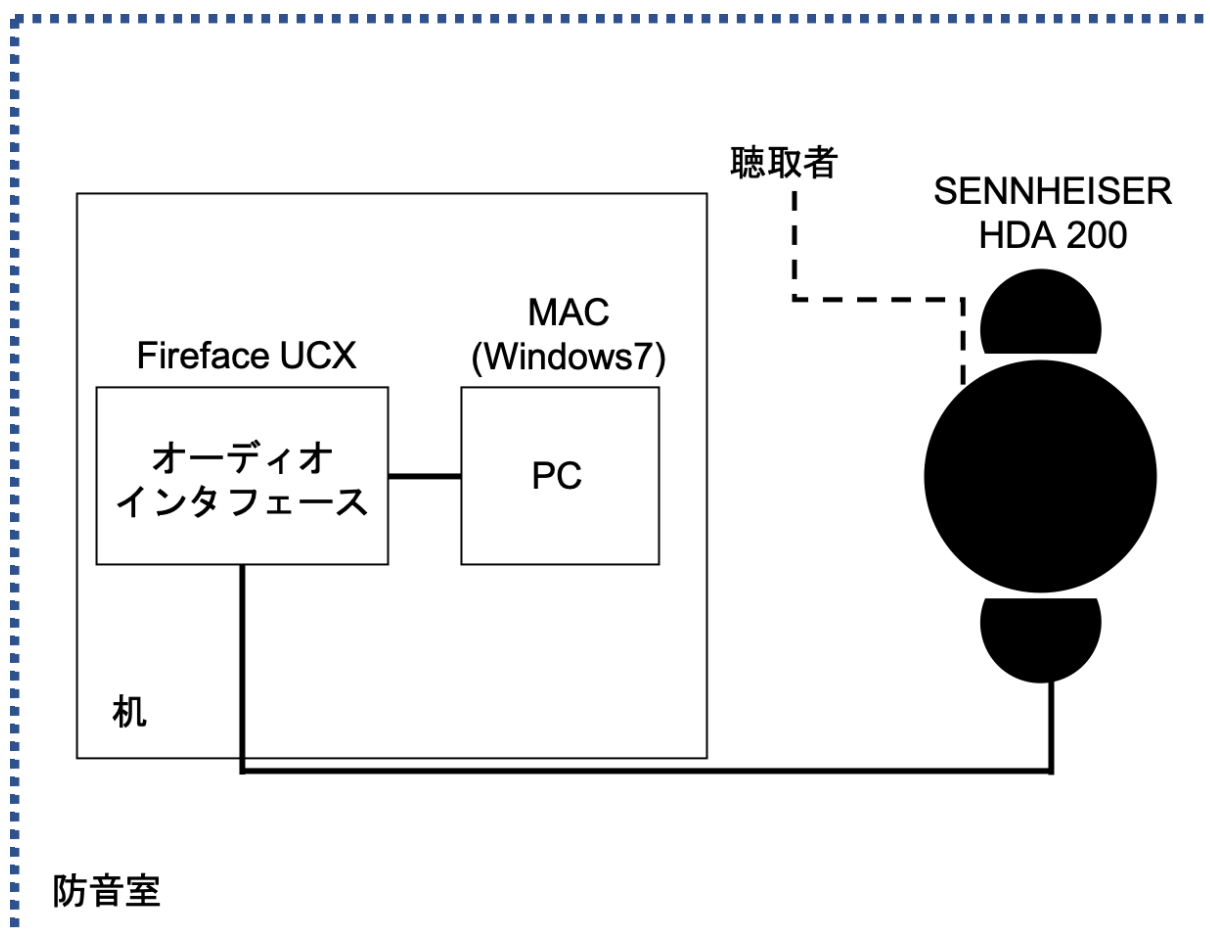


図 4.3: 実験環境

4.4 実験結果

図 4.4 に男性話者の実験結果を，図 4.5 に女性話者の実験結果を示す．横軸は話者の実年齢であり，縦軸は知覚年齢を示す．図 4.4 において図中の丸は各実年齢に対する知覚年齢の平均値の結果を示す．実線は結果に対する回帰直線を示す．また，図 4.5 において図中の $\times$ は各実年齢に対する知覚年齢の平均値の結果を示す．図中のエラーバーは聴取者による推定値のばらつきを示している．実線は結果に対する回帰直線を示す．図 4.4，図 4.5 の結果から，聴取者は話者の年齢を正確に推定していることが分かる．また，男性話者の場合も女性話者の場合も子供の音声はばらつきが大人に比べ小さくなっているのが見られる．大人，高齢者の音声の結果を見ると実年齢が 30 代から 50 代の音声が他の年齢の音声に比べばらつきが多いように見える．回帰直線を見ると，女声の場合は実年齢に比べ知覚年齢が高く推定されていることを示している．相関値を見ると男性の場合も女性の場合も $r^2 = 0.92$ と高い相関を示している．

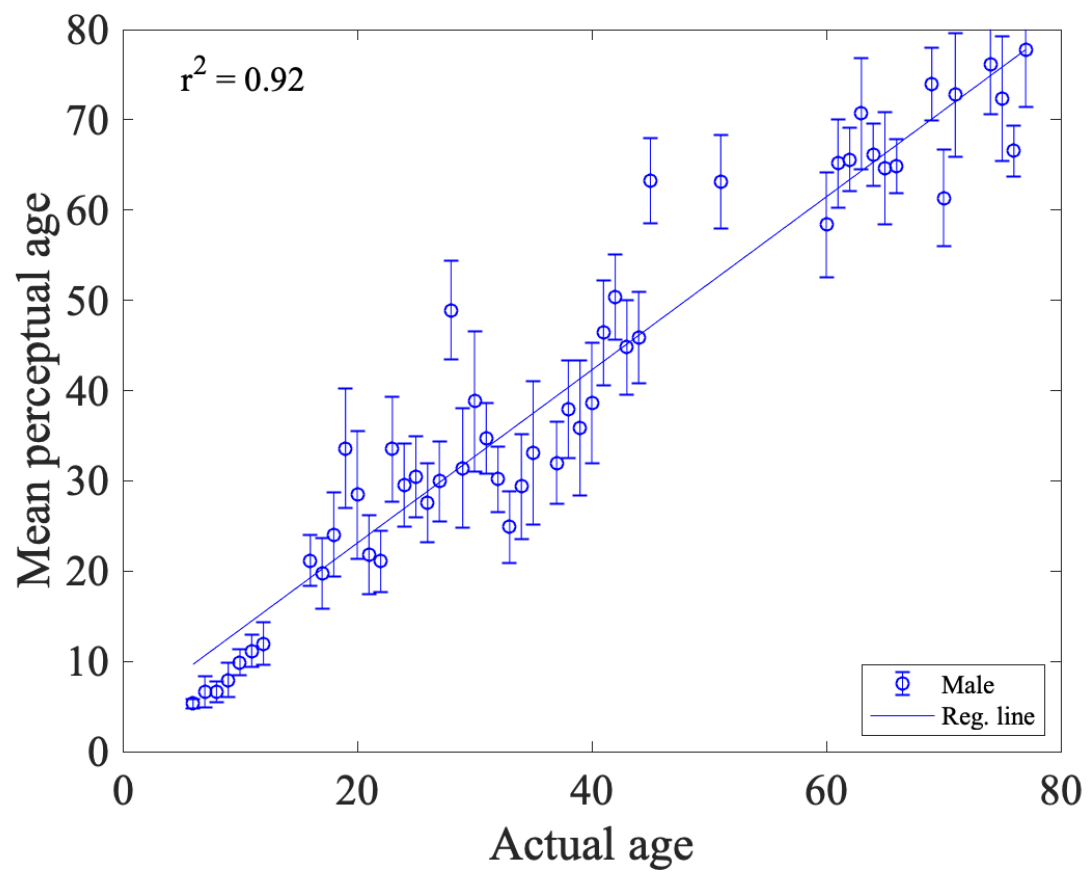


図 4.4: 実年齢と知覚年齢の関係 (男声)

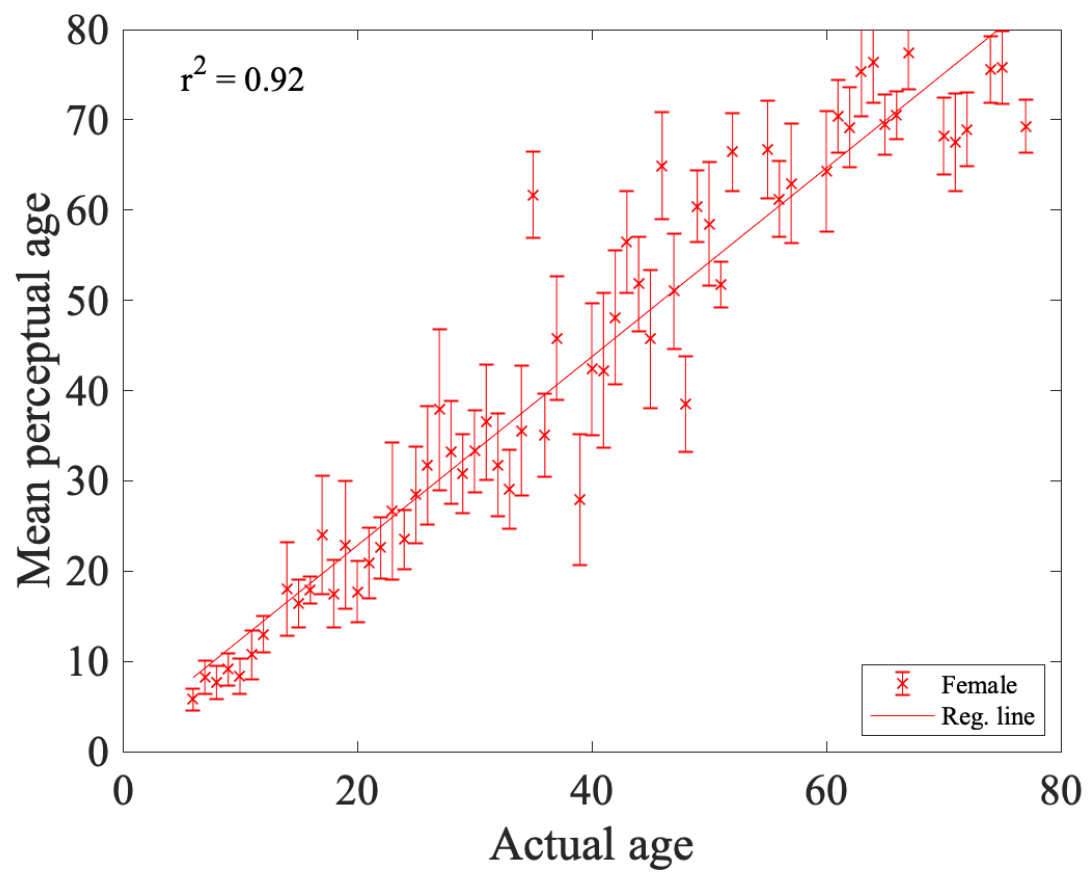


図 4.5: 実年齢と知覚年齢の関係 (女声)

4.5 考察

今回得られた結果から、男性話者の場合、20代の音声は実年齢より知覚年齢が高くなり、40代の音声は実年齢が知覚年齢よりも低くなっていることが見える。木戸ら報告によると、知覚年齢は35歳を境とし、実年齢が35歳より下の場合は老けて、35歳より上の場合は若く聞こえる傾向にある[30]。今回の結果は、これらの結果に類似した傾向がみられる。しかし、聴取者の知覚年齢にばらつきも見えるので、個人差がある可能性がある。また、60歳以降の音声では、知覚年齢が低く見積もられている音声が多く見える。この結果は、水町の高齢者音声に対する知覚年齢を求めた結果と類似した結果が得られていると考えられる[13]。そして、男性音声の場合も、女性音声の場合も、実年齢と知覚年齢の相関が高いことから、人は年齢をある程度正確に知覚できていることが考えられる。しかし、実年齢と知覚年齢には差があるため、知覚年齢による検討も必要であると考えられる。

第5章 年齢知覚に関わる特徴の分析

5.1 分析方法

先行研究によって男性の場合はスペクトル傾斜が、女性の場合は基本周波数 F0 が特に年齢との関係があることが明らかされた。そこで、まず聴取実験で得られた各音声の知覚年齢と音響特徴の関係を調査した。ここでは、発話内容のバラつきを考慮して、話者当たり 10 種類の音声刺激に対する音響特徴の平均値を用いた。F0 は、有声区間に対して STRAIGHT TENPO を利用して求め、その結果を時間平均し、音声の F0 平均を求めた [29]。また、スペクトル傾斜は、フレーム長 10 ms、フレームシフト 5 ms、ハニング窓を窓関数として短時間パワースペクトルを求め、オクターブ周波数上のスペクトル傾斜を最小二乗法により求めた。次に、聴取実験で得られた各音声の知覚年齢と音質評価指標（ラフネス、変動強度、シャープネス）との関係を求めた。実験刺激に対する音質評価指標は OROS 社の Sound Quality Lite を用いて算出した。ここでは、発話内容のバラつきを考慮して、話者当たり 10 種類の音声刺激に対する音響特徴と音質評価指標の平均値を用いた。

5.2 知覚年齢と音響特徴の関係

図 5.1 に男性話者の知覚年齢とスペクトル傾斜の関係を、図 5.2 に女性話者の知覚年齢とスペクトル傾斜の関係を示す。横軸は話者の知覚年齢であり、縦軸はスペクトル傾斜を示す。図 5.1 において図中の ○ は各知覚年齢に対するスペクトル傾斜の平均値の結果を示す。図中のエラーバーは発話内容による特徴量のばらつきを示している。実線は結果に対する回帰直線を示す。また、図 5.2 において図中の × は各知覚年齢に対する音響特徴の平均値の結果を示す。実線は結果に対する回帰直線を示す。

図 5.1 図 5.2 の結果から年齢全体を見ると男性の音声の場合も女性の音声の場合も年齢が上がるとともにスペクトル傾斜が増加する傾向が得られた。男性話者の場合、子供と高齢者の音声において、スペクトル傾斜のばらつきが大きい。女性の場合は、どの音声についてもばらつきが大きい。

図 5.3 に男性話者の知覚年齢と F0 の関係を、図 5.4 に女性話者の知覚年齢と F0 の関係を示す。図中の記号は図 5.1、図 5.2 と同様である。

図 5.3 図 5.4 の結果から年齢全体を見ると男性の音声の場合も女性の音声の場合も知覚年齢の増加に伴って F0 が減少している傾向が得られた。F0 は、どの年齢においてもばら

つきが小さくなっていることが分かる。また、女性の音声の F0 は緩やかに減少しているように見えるのに対して、男性の音声では、10 代から 20 代で急激に減少しているように見える。

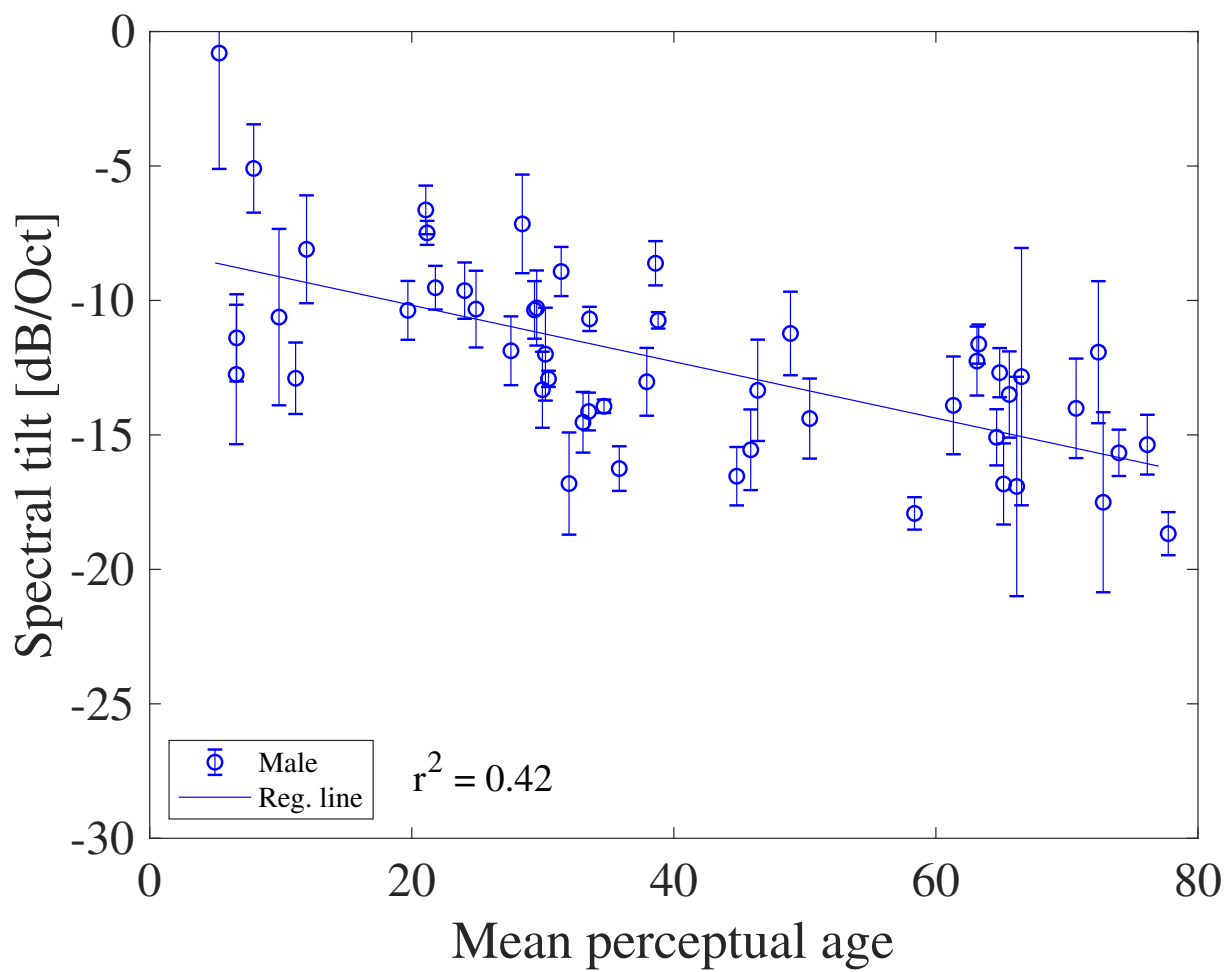


図 5.1: 知覚年齢とスペクトル傾斜の関係 (男声)

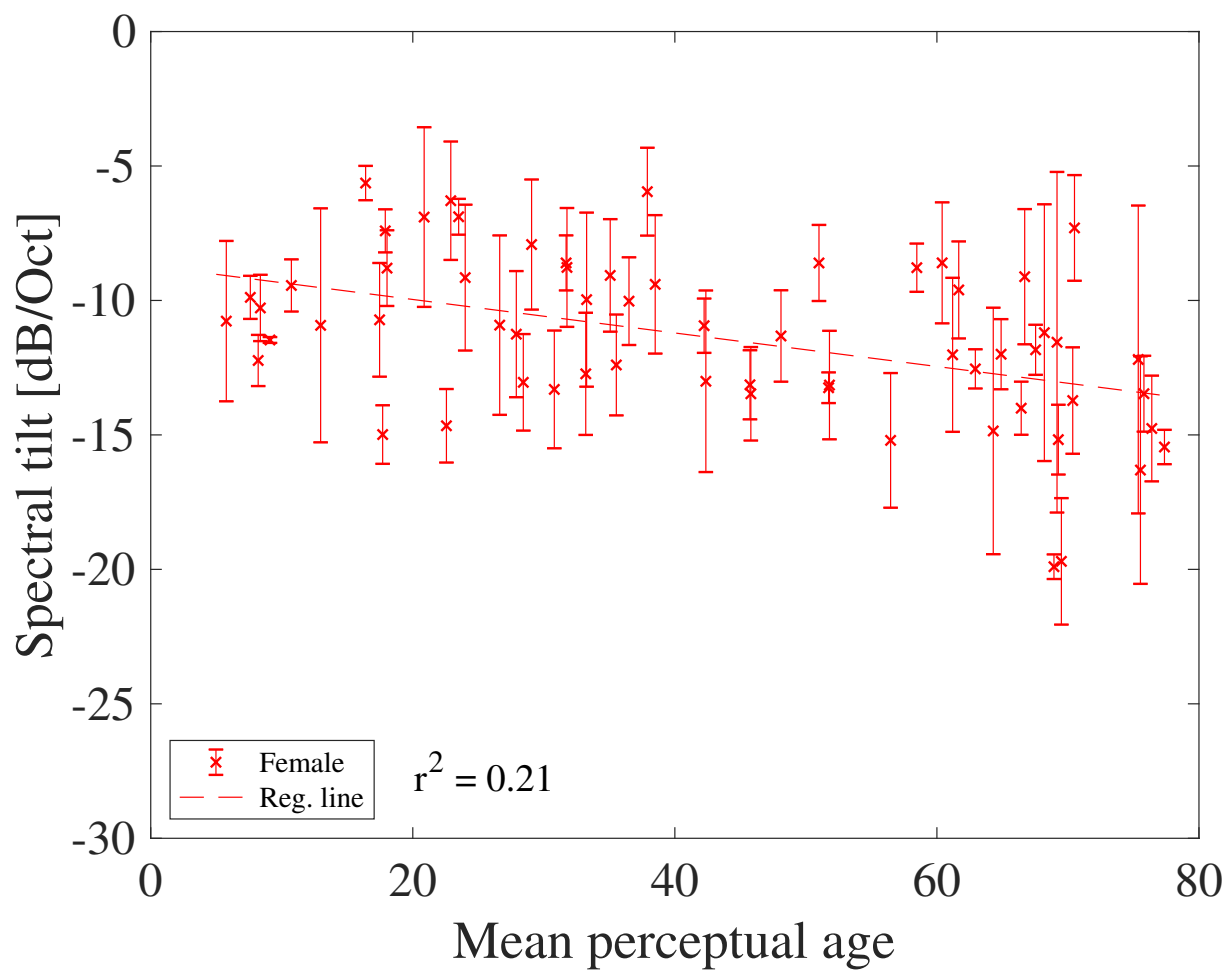


図 5.2: 知覚年齢とスペクトル傾斜の関係 (女声)

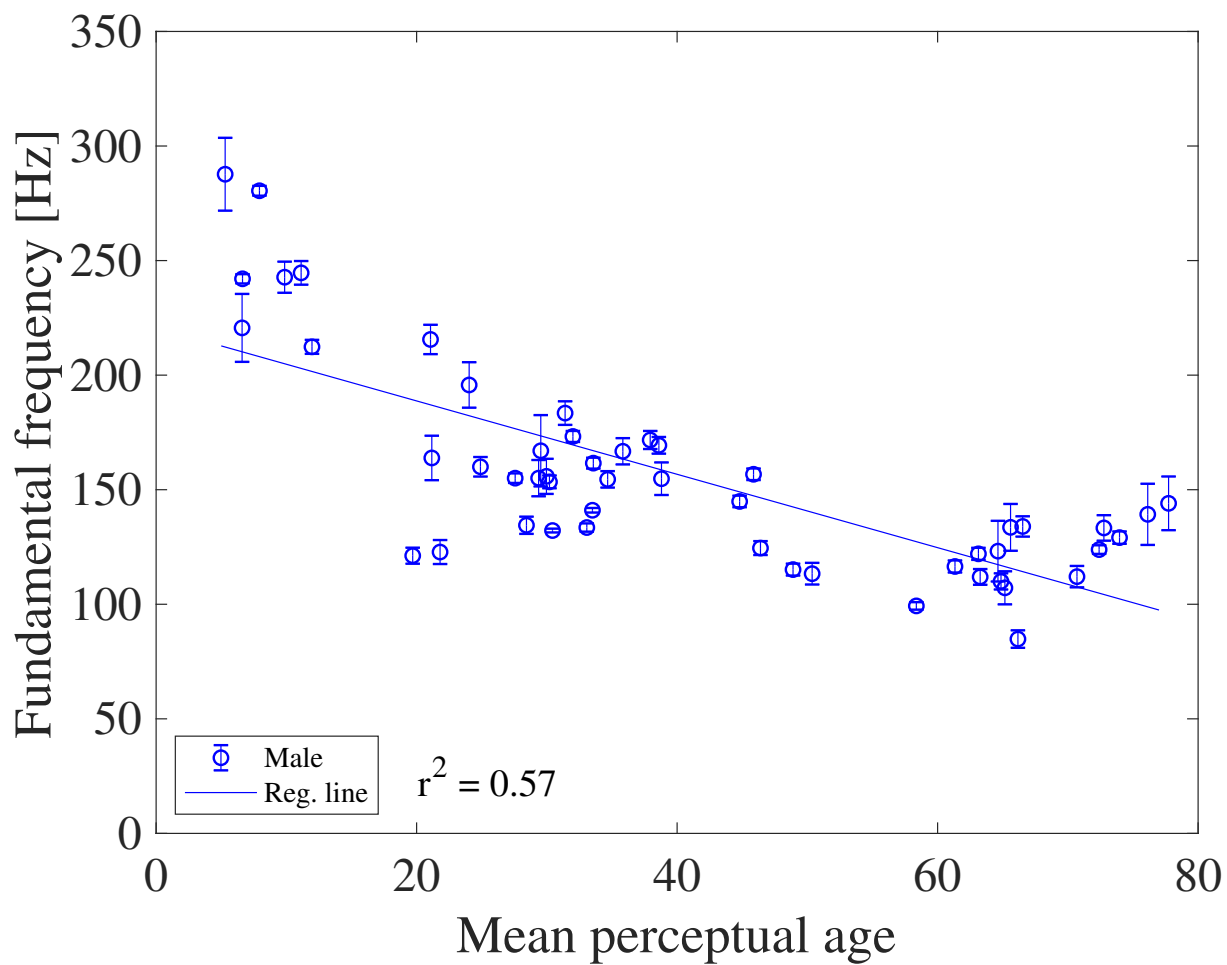


図 5.3: 知覚年齢と F0 の関係 (男声)

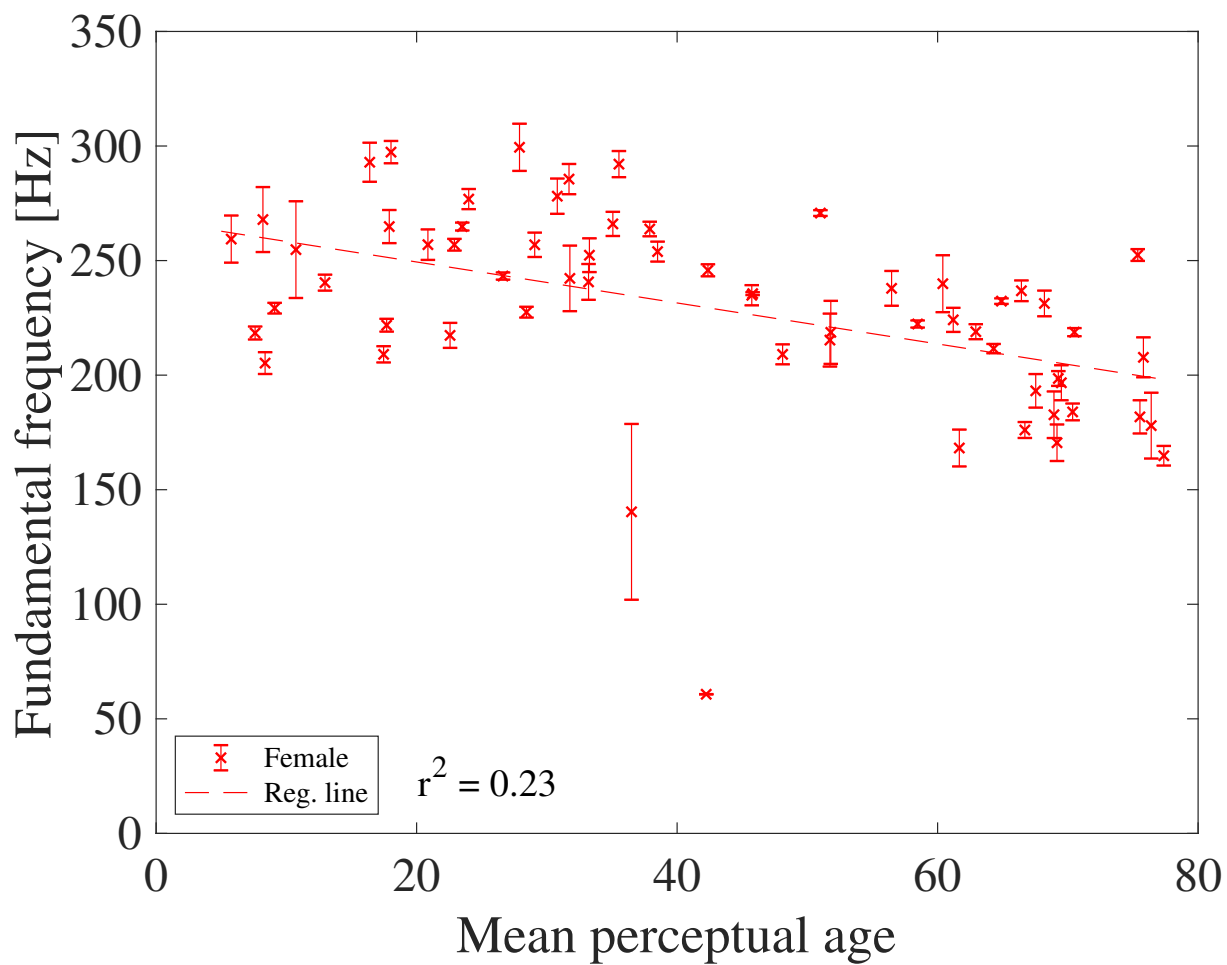


図 5.4: 知覚年齢と F0 の関係 (女声)

5.3 知覚年齢と音質評価指標の関係

図 5.5 に男性話者の知覚年齢とラフネスの関係を、図 5.6 に女性話者の知覚年齢とラフネスの関係を示す。横軸は話者の知覚年齢であり、縦軸はラフネスを示す。図 5.5 において図中の ○ は各知覚年齢に対するラフネスの平均値の結果を示す。図中のエラーバーは発話内容による特徴量のばらつきを示している。実線は結果に対する回帰直線を示す。また、図 5.6 において図中の × は各知覚年齢に対するラフネスの平均値の結果を示す。実線は結果に対する回帰直線を示す。

図 5.5 図 5.6 の結果から年齢全体を見ると男性の音声の場合は知覚年齢が増加するとともにラフネスが上昇する傾向が見られる。女性の音声の結果を見ると、知覚年齢による上昇傾向の下降傾向も見られなかった。しかし、20 歳から 60 歳の音声は、子供や高齢者の音声に比べてラフネスの値が小さくなっているようにも見える。

図 5.7 に男性話者の知覚年齢と変動強度の関係を、図 5.8 に女性話者の知覚年齢と変動強度の関係を示す。図中の記号は図 5.5、図 5.6 と同様である。

図 5.7 図 5.8 の結果から年齢全体を見ると男性の音声の場合は知覚年齢が増加するとともに変動強度が減少する傾向が見られる。処世の音声の場合も回帰直線を見ると減少傾向が見られる。しかし、変動強度は他の指標に比べて、ばらつきが大きくなっていることも見える。

図 5.9 に男性話者の知覚年齢とシャープネスの関係を、図 5.10 に女性話者の知覚年齢とシャープネスの関係を示す。図中の記号は図 5.5、図 5.6 と同様である。

図 5.9 図 5.10 の結果から年齢全体を見ると男性の音声の場合は年齢全体で見ると知覚年齢に対するシャープネスに傾向は見られなかった。女性の音声の場合は、多少の減少傾向が見えるもののばらつきが大きくなっている。

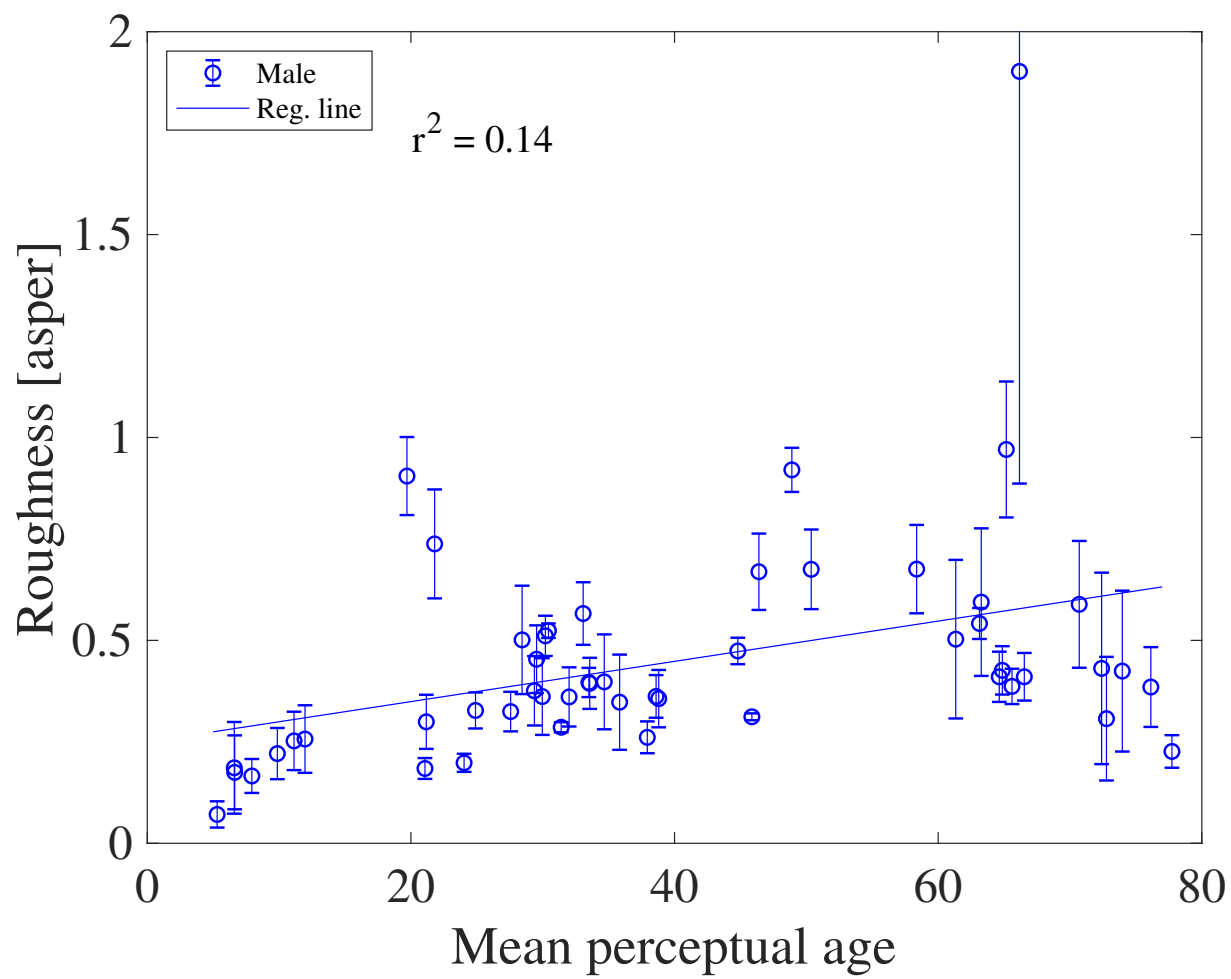


図 5.5: 知覚年齢とラフネスの関係 (男声)

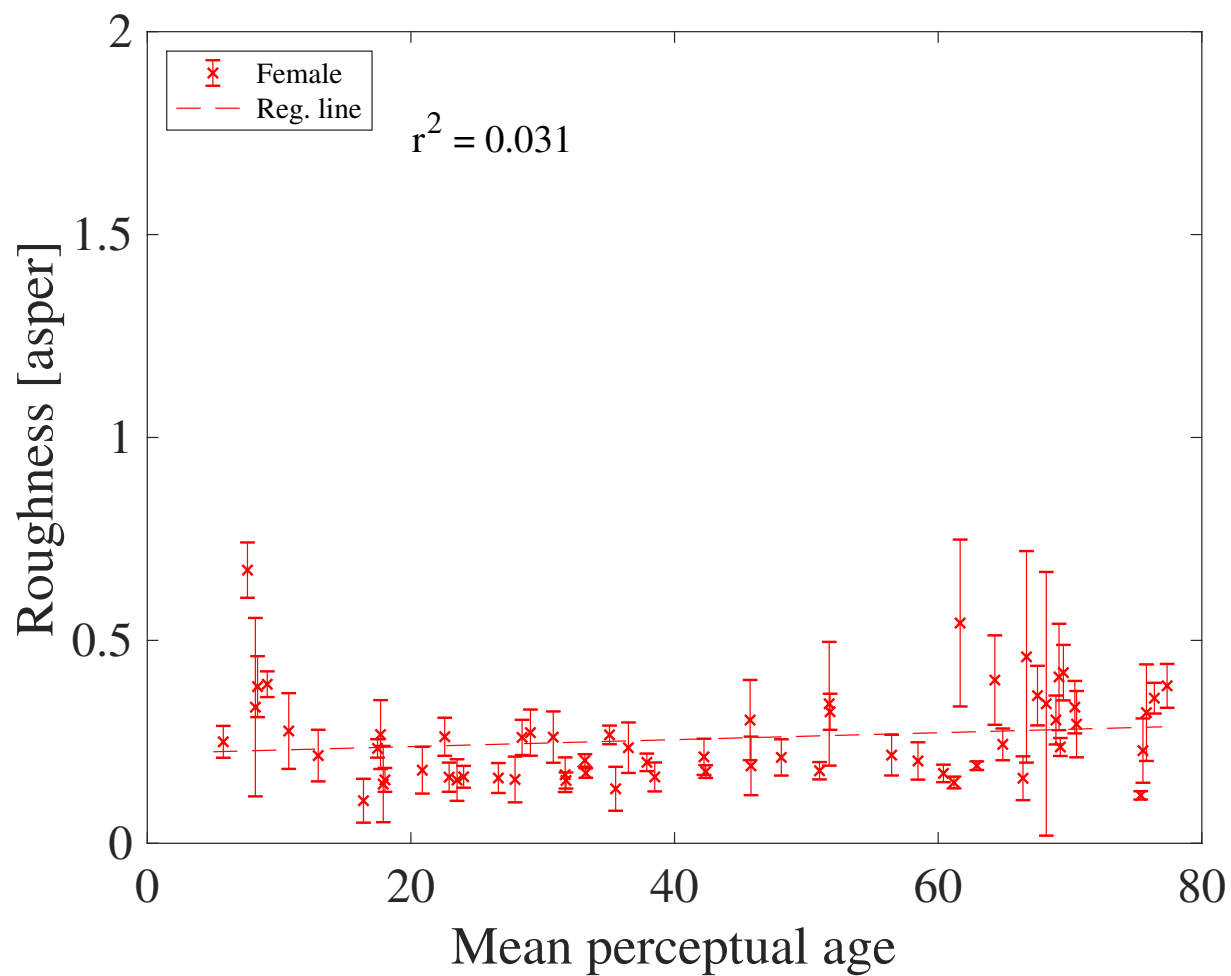


図 5.6: 知覚年齢とラフネスの関係 (女声)

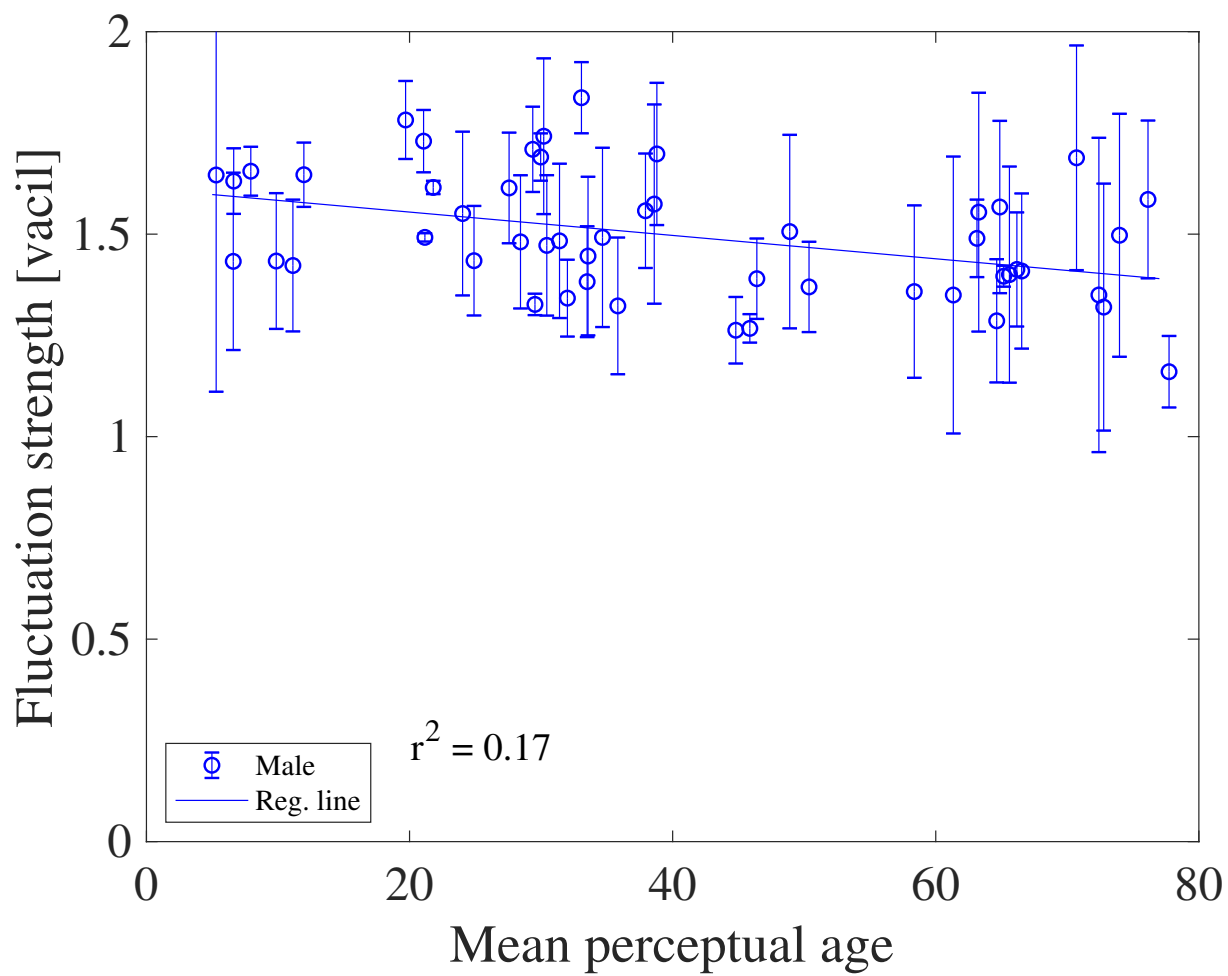


図 5.7: 知覚年齢と変動強度の関係 (男声)

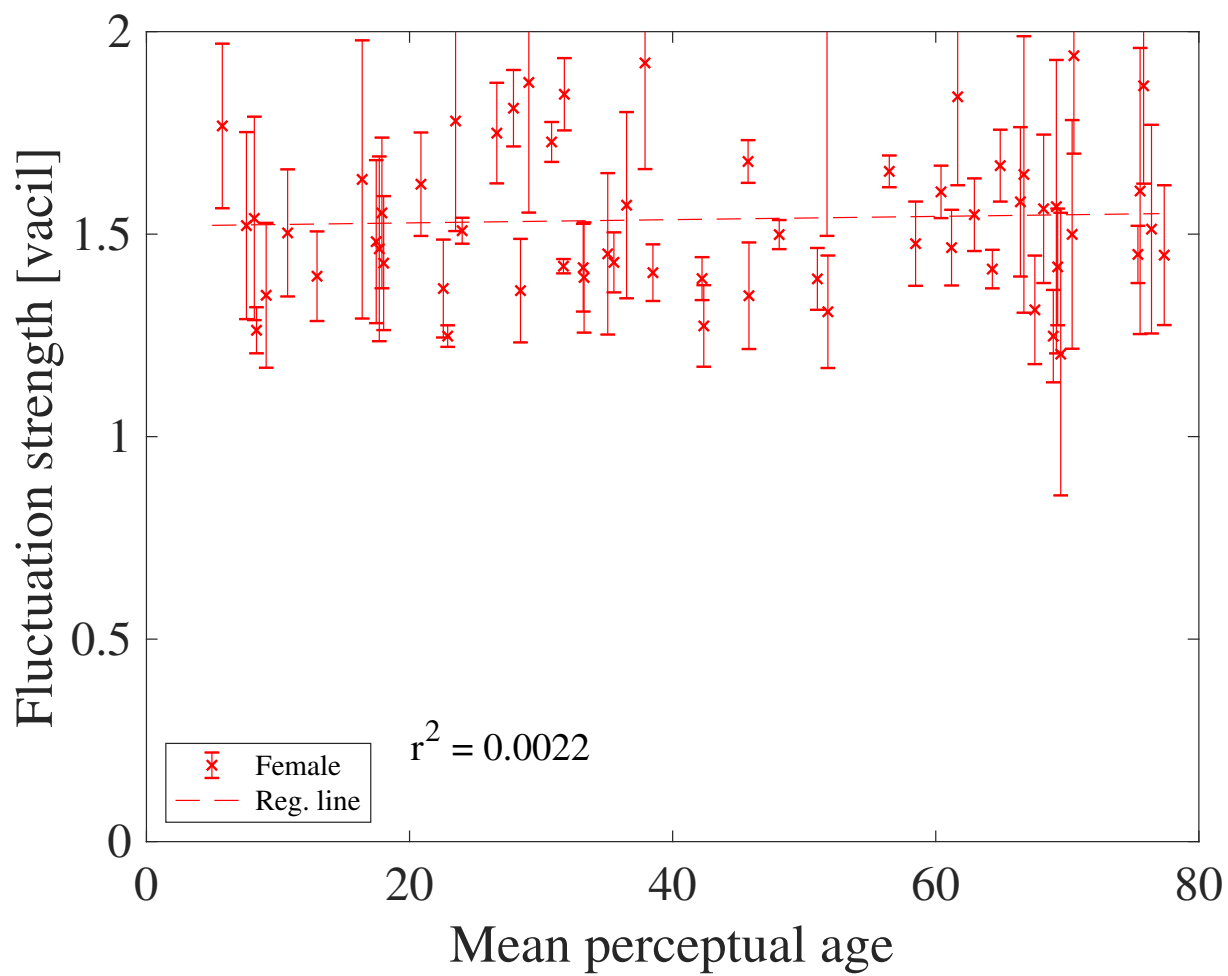


図 5.8: 知覚年齢と変動強度の関係 (女声)

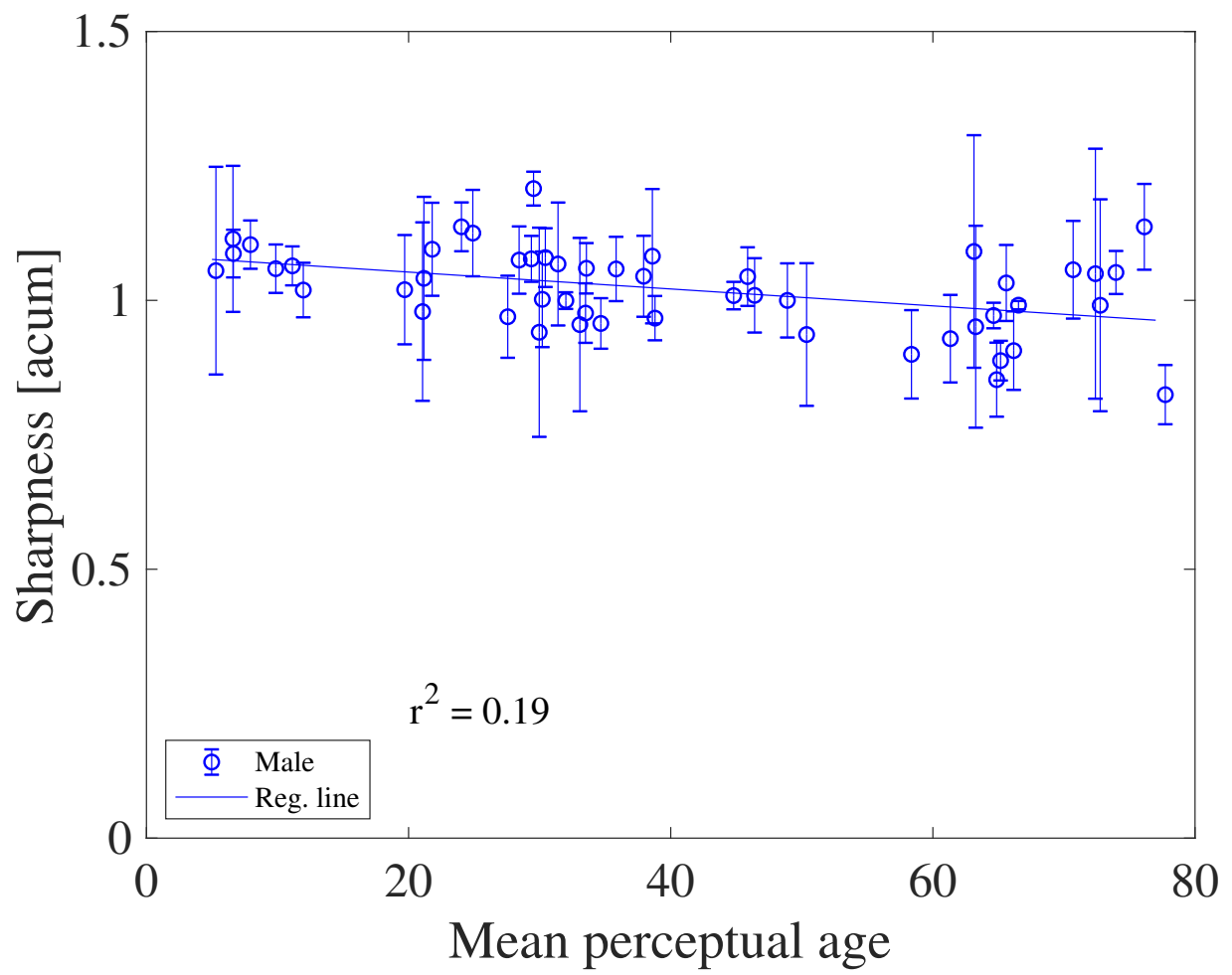


図 5.9: 知覚年齢とシャープネスの関係 (男声)

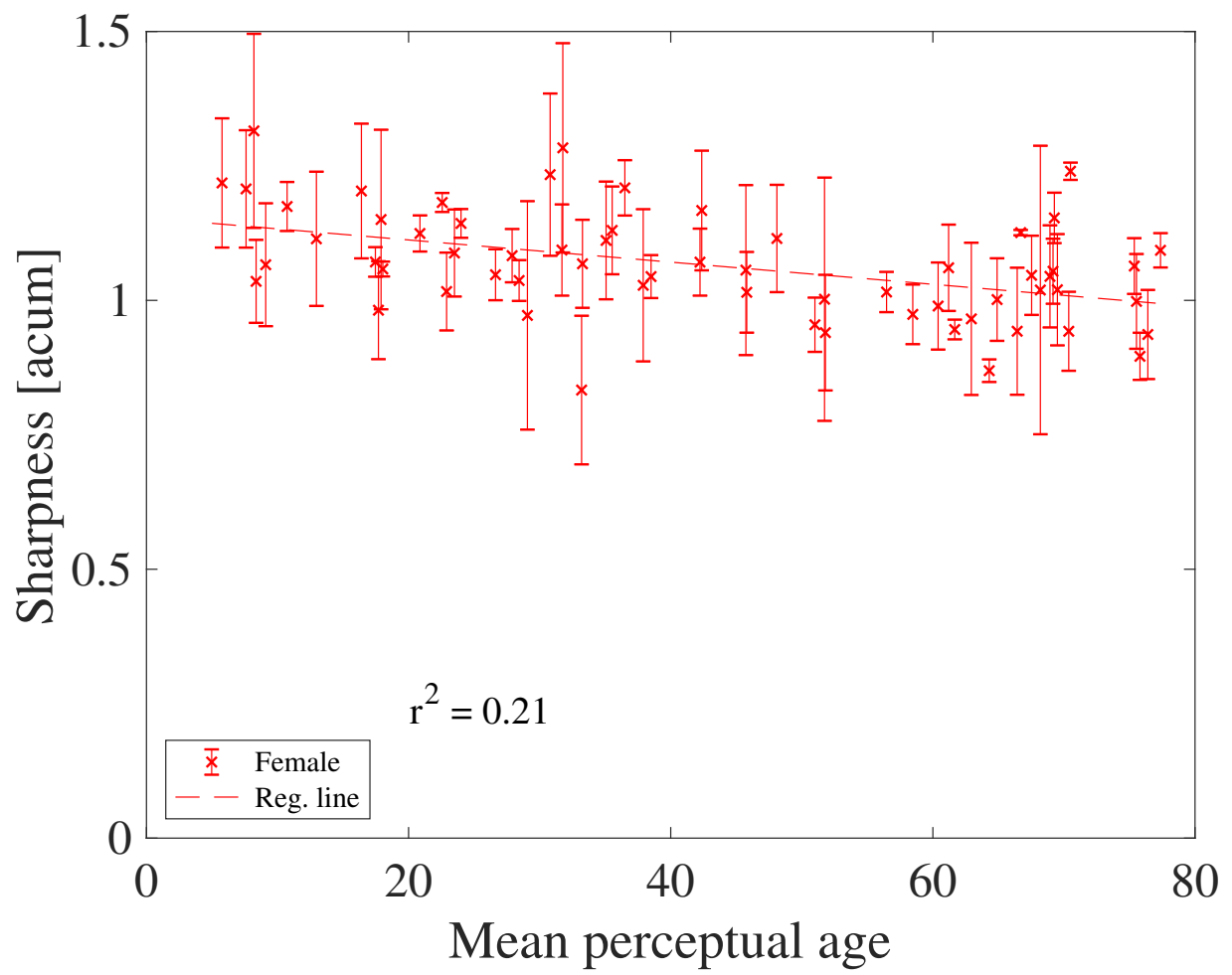


図 5.10: 知覚年齢とシャープネスの関係 (女声)

5.4 音響特徴と音質評価指標の結果の比較

知覚年齢と音響特徴の関係を見た場合と知覚年齢と音質評価指標の関係を見た場合を比較すると、子供、大人、高齢者の全体の知覚年齢に対して最も相関が高かったのは男性の場合は、F0 の $r^2 = 0.57$ であった。次に高かったのは、スペクトル傾斜の $r^2 = 0.42$ であった。女性の場合は、F0 の $r^2 = 0.23$ であった。次に高かったのは、スペクトル傾斜とシャープネスの $r^2 = 0.21$ であった。しかし、スペクトル傾斜は、ばらつきが激しいので他の特徴量が関係している可能性がある。

5.5 考察

知覚年齢とスペクトル傾斜、ならびに知覚年齢と F0 の関係を見た結果より、男性音声の知覚年齢とスペクトル傾斜に傾向が見られたことから知覚年齢に関しては従来知られている音響特徴で、今回のデータを説明できている可能性が考えられる。しかし、知覚年齢に対する相関の値は低いため知覚年齢に関しては聴覚の印象に係るところで求める必要があると考えられる。知覚年齢と音質評価指標との関係の結果を見ると、年齢全体を説明しうる指標は見つからなかったため音質評価指標では説明ができないのではないかと考えられる。シャープネスはスペクトル傾斜にも関係する指標であるので、スペクトル傾斜が急になる、つまり高域の周波数成分が減るのであればシャープネスが減少する傾向が見られると考えられるが、今回の結果にはそのような傾向が見られなかった。また、水町らの検討によると高齢者の音声では、高域の周波数成分が減り、パワースペクトルのダイナミックレンジが縮小すると報告されているが、今回の結果では知覚年齢の上昇によってスペクトル傾斜が急になってることを考えると、先行研究とは異なる結果が得られていると考えられる。

第6章 全体考察

実年齢と知覚年齢を見ると実年齢と知覚年齢の間の相関が高く、聴取者は話者の年齢を正確に推定できている可能性が示された。本研究では、実験の際に聴取者に話者の性別を教示してから実験を行っているため、実年齢と知覚年齢の値がかなり近い値になっている可能性がある。そのため、実験の際に性別を教示せずに行った場合に分析結果が変動がある可能性がある。その後の分析で、スペクトル傾斜と F0 が知覚年齢と相関が高いという結果が得られた。スペクトル傾斜や F0 は先行研究で実年齢との関係が示唆されている。このことから、実年齢と知覚年齢が近い値であったことが結果に作用していることが考えられる。次に、音質評価指標による結果を考える。まず、シャープネスが知覚年齢に対して下降傾向を示した結果は、シャープネスはスペクトル傾斜に関連する指標であることが挙げられる。本研究ではスペクトル傾斜が知覚年齢が上がるに連れて傾斜が急になるという結果が得られた。甲高さの知覚は傾斜が正の勾配、フラット、負の勾配の順に高くなるためシャープネスの値が知覚年齢が上昇するに連れて下降傾向を示したと考えられる。また、人は高齢者の音声を聞いたときに、ざらつき感を感じると考えられる。そのため、分析前には、知覚年齢の上昇によりラフネスもともに上昇していき、高齢者に感じられる音声はラフネスが急上昇すると予測していた。感覚的な考え方でだけでなく、人は高齢になると声帯が硬化し、振動しづらくなると考えられる。この影響から、音声にざらつき感が感じられるかとも考えていた。しかし、上記のような急激な変化は今回の分析からは見られなかった。男性話者の結果を見るとラフネスが上昇しているように見えるが、予測していたような変化ではないと考えられる。しかし、ラフネスは 1 asper のときに人が最もざらつき感を感じる指標であるということを考えると、本研究の結果は、男性の知覚年齢が最年少のときにラフネスの値が 0.2 asper であるのに対して知覚年齢が成人、高齢者の年齢になると 0.5 asper 程に上昇している。これを加味して考えるとラフネスは知覚年齢に対して相関としては低いものの、年齢の知覚に関わっている可能性は十分にあると考えられる。実際に、聴取実験を行ったあとに、内観報告を行ったところ、ざらつき感から知覚年齢を答えたとの報告もあり、音質評価指標の内特にざらつき感を表すラフネスは、知覚年齢に関係していることを否定し得ないとも考えられる。

第7章 結論

7.1 本研究で明らかになったこと

本研究の目的は、聴覚的要因として音色知覚に係る音質評価指標（ラフネス、変動強度、シャープネス）を用いて、人の年齢知覚に関する知見を深めることを目的として分析を行った。そのために、まず話者の実年齢に対する知覚年齢を求めた。その結果、聴取者は話者の年齢を正確に知覚できているという結果が得られた。その後、従来研究で実年齢との関係が示唆されたスペクトル傾斜及びF0と知覚年齢との関係を確認した上で、音質評価指標を用いて知覚年齢との関係性を求める分析を行った。その結果、音質評価指標は本研究で利用したデータの年齢全体を説明し得ないという結果が得られた。しかし、知覚年齢にばらつきがあることを考え、全データの内子供、大人、高齢者から標準偏差が小さいデータをそれぞれ3名のデータ、4名のデータ、3名のデータを抽出し、再分析を行った。この結果から、変動強度やシャープネスは、年齢全体に対する結果と同じく年齢による傾向は見られなかったが、ラフネスの関係を否定し得ないという結果が得られた。

人は従来研究で明らかにされたF0などの他に音声のざらつき感を年齢知覚の指標にしている可能性があることが示された。

7.2 残された課題

・知覚年齢を求める実験の改善

本研究では、子供、大人、高齢者に対応する3種類の音声データベースを統合して聴取実験を行った。しかし、4章でも記述した通り子供の音声のみが発話内容が違ったために、実験前に発話内容によらないように注意したものの、年齢知覚の頼りにした可能性は捨てきれない。そのため、年齢知覚についてより知見を深めるのであれば、子供から高齢者までの広い年齢帯を含む同発話内容のデータベースが必要であると考えられる。・音色・個人性知覚メカニズムの解明

本研究では、覚的要因として音色知覚に係る音質評価指標の内のラフネス、変動強度、シャープネスを用いた。ラフネスにおいては年齢知覚の手がかりになっている可能性が示唆された。この結果から、年齢知覚には音声生成系による分析の他に聴覚的要因も考慮しなければいけないことが考えられる。よって本研究で用いた特徴量の他の年齢知覚に関わる特徴量を調べることで、聴知覚メカニズム、個人性知覚メカニズムの解明につながることを考えられる。

参考文献

- [1] 粕谷英樹, “声質の伝える情報とその関連量,” 日本音響学会誌, vol. 68, no. 10, pp. 520-526, 2012.
- [2] 森大毅, 前川喜久雄, 粕谷英樹, 音響サイエンスシリーズ 12 音声は何を伝えているかー感情・パラ言語情報・個性の科学ー, コロナ社, 2014.
- [3] Kasuya, H., Yoshida, H., Ebihara, S. and Mori, H., “Longitudinal changes of selected voice source parameters,” Proc. interspeech2010, pp. 2570 – 2573, 2010.
- [4] 西尾正輝, 田中康博, 新美成二, “加齢に伴う音声の変化ー音響学的手法を用いた解析ー,” 音声言語学 vol. 50, pp. 6-13, 2009.
- [5] 栗田茂二郎, “声帯の成長, 発達と老化ーとくに層構造の加齢的变化ー,” 音声言語医学 vol. 29, No. 2, pp. 185-193, 1988.
- [6] 平野 実, 栗田茂二郎, 藤雄一, “声帯の成長, 発達と老化ーとくに声帯粘膜の変化についてー,” 耳鼻臨床 vol. 74, No. 8, pp. 1791-1802, 1981.
- [7] 平野 実, 坂口伸治, 栗田茂二郎, “声帯粘膜の老化,” 喉頭 1, pp. 31-35, 1989.
- [8] Fitch T.M., Giedd J., “Morphology and development of the human vocal tract: A study using magnetic resonance imaging,” J. Acoust. Soc. Am., 106, 1511 – 1522, 1999.
- [9] 麦谷綾子, 廣谷定男, “子供の声道発達と音声の特性変化,” 日本音響学会誌, vol. 68, no. 5, pp. 234-240, 2012.
- [10] 今泉, “声質の計量心理学的評価,” 日本音響学会誌, vol. 42, no. 10, pp. 828-833, 1986.
- [11] 日本音声言語医学会, 新編声の検査法, 医歯薬出版会社, 2009.
- [12] 廣瀬, 音声障害の臨床, インテルナ出版, 1998.
- [13] 水町光徳, “加齢による聴力低下と発話変化の相互作用,” 日本音響学会誌 vol. 73, no. 5, pp. 279-302, 2017.

- [14] 堀紗矢香, 田中章浩, “音声の基本周波数と抑揚が知覚年齢に及ぼす効果,” 日本認知科学会, pp. 694-697, 2014.
- [15] 重野純, “感情音声による年齢知覚,” 日本心理学会第 81 回大会 9. 認知, 2017.
- [16] 小林和弘, 土井啓成, 戸田智基, 中野倫靖, 後藤真考, ニュービッググラム, サクリアニサクテイ, 中村哲, “知覚年齢に沿った歌声声質制御のための音響特徴量の調査,” 情報処理学会報告書, no. 44, pp. 1-6, 2013.
- [17] 小林和弘, 戸田智基, 中野倫靖, 後藤真考, Graham Neubig, “統計的手法に基づく歌声の知覚年齢制御,” 電子情報通信学会技術研究報告, pp. 321-326, 2014.
- [18] 笠原洋子, 巖島行雄, “声の認識-声の様態が年齢推定に与える影響,” 法と心理, pp. 71-84, 2007.
- [19] 峯松信明 “音声に含まれる様々な個人性の自動推定,” 日本音響学会誌 vol. 71, no. 9, pp. 490-497, 2015.
- [20] 山内景太, 峯松信明, 廣瀬啓吉, “PUI を目的とした音声情報からの知覚的年齢の自動推定,” 情報処理学会第 65 回全国大会講演論文集, pp. 235-236, 2003.
- [21] 峯松信明, 廣瀬啓吉, 関口真理子, “話者認識技術を利用した主観的高齢話者の同定とそれに基づく主観的年代の推定,” 情報処理学会論文誌 vol. 43, no. 7, pp. 2186-2196, 2002.
- [22] 山内景太, 峯松信明, 廣瀬啓吉, “話者認識技術を利用した主観的高齢話者の同定とそれに基づく主観的年代の推定,” 電子情報通信学会技術研究報告, pp. 43-48, 2003.
- [23] 岩宮眞一郎, 音響サイエンスシリーズ 1 音色の感性学—音色・音質の評価と創造, コロナ社, 2010.
- [24] Fastl H., “Fluctuation strength and temporal masking patterns of amplitude-modulated broadband noise,” Hearing Research, 8, pp. 59-69, 1982.
- [25] Fastl, H., Zwicker E., “Psychoacoustics : Facts and models, third edition” Springer, pp. 257-264, 2006.
- [26] 新聞記事読み上げ高齢者音声コーパス <http://research.nii.ac.jp/src/S-JNAS.html> (2月3日アクセス)
- [27] ATR 多数話者データベース <http://www.atr-p.com/products/sdb.html> (2月3日アクセス)
- [28] 子供の声データベース <http://research.nii.ac.jp/src/CLAIR-VCV.html> (2月3日アクセス)

- [29] Kawahara, H., Katayose, H., and de Cheveigne, A., Patterson, R., “Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity,” Proc. Eurospeech99, pp.2781-2784, Sept. 1999.
- [30] 木戸博, “音声聴取における声年齢についての一考察,” 音講論, pp. 491-492, 2004.

研究業績一覧

国際会議

1. HATAKEYAMA, T., UNOKI, M., “Study on the method for estimating perceptual age using sound quality metrics,” 2019 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP’19), 2019.

国内発表

1. 畠山達也, 鵜木祐史, “音声の知覚年齢と音質評価指標の関係,” 平成 30 年度 電気関係学会北陸支部連合大会, G-3, 2018.
2. 畠山達也, 鵜木祐史, “音質評価指標を用いた年齢知覚に関する検討,” , EA2018-74, 65–70, 2018.

謝辞

研究活動から私生活，社会の一般常識にいたるまで，厳しくも非常に丁寧な指導をしていただいた，主指導教員である鵜木祐史教授には，深謝の意を表する．また，研究室会議やミーティングなどの場において，数多くの助言を頂いた赤木正人教授に深謝の意を表す．研究で行き詰まったときに相談に乗って頂いた，鵜木・赤木研究室のメンバーの皆様に感謝致します．本学での研究生生活を支え，温かく見守ってくれた両親に心から感謝致します．