

音質評価指標を用いた年齢知覚に関する検討

畠山 達也 鵜木 祐史

北陸先端科学技術大学院大学 先端科学技術研究科

〒923-1292 石川県能美市旭台 1-1

E-mail: {tatsuya_hatakeyama, unoki}@jaist.ac.jp

あらまし 音声の年齢知覚は主に音声生成メカニズムの観点から検討されてきたが、未だにその年齢知覚メカニズムは明らかになっていない。本稿では、音声の年齢知覚が聴覚の変調知覚に関係しているものと予想し、音声の知覚年齢と聴覚的要因（音質評価指標）との関係を調査する。はじめに、若年者と高齢者の音声データベースを利用して知覚年齢実験を行った。その後で、音声の知覚年齢と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との関係を調べた。その結果、実年齢と知覚年齢の関係から、知覚年齢は話者の年齢が上がるにつれてバラつきが見られることがわかった。特に成人音声の場合、35歳を起点に実年齢と知覚年齢の間に老若の逆転がみられることがわかった。また、知覚年齢と音質評価指標の関係から、男性音声の場合は、音声のラフネスと変動強度が、女性音声の場合はラフネスやシャープネス変動強度が知覚年齢を説明できることがわかった。

キーワード 年齢知覚, 音質評価指標, ラフネス, 変動強度, シャープネス

Study on age perception using sound quality metrics

Tatsuya HATAKEYAMA, Masashi UNOKI

Graduate School of Advanced Science and Technology, Japan Advanced Institute of Science and Technology

1-1 Asahidai, Nomi, Ishikawa, 923-1292 Japan

E-mail: {Tatsuya_hatakeyama, unoki}@jaist.ac.jp

Abstract The age perception of human voice has been studied mainly from the viewpoint of speech production mechanism. However, the age perception has not been revealed yet. In this paper, we predict that acoustical features related to auditory modulation perception can be used to take account for the age perception. We then investigate a relationship between the age perception and auditory modulation perception by carrying out experiments for age perception. In the experiments, two kinds of speech corpus (children and elderly voices) were used. Then, we investigated relationships between perceptual age and auditory factors such as roughness, fluctuation strength, and sharpness. As results, it was found that the deviation between actual age and perceived age increases as the age of the speaker increases. In particular, it was found that the perceived ages of children and the elderly people were reversed at the actual age of 35. In addition, from the relationship between perceived age and sound quality metrics, it was found that roughness and fluctuation strength can be used in the case of male speech to account for the corresponded perceptual age, and that roughness, fluctuation strength, and sharpness can be used in the case of female speech to account for the corresponded perceptual age.

Keywords Age perception, Sound quality metrics, Roughness, Fluctuation strength, Sharpness

1. はじめに

ヒトは、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。しかし、どのようなメカニズムで年齢を知覚しているかまでは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、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、これまでの検討結果から、音声生成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の検討が行われてき

た[1][2]. 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ることも分かった.

「ことばの鎖」を振りどころに, 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, 高齢者においては, 加齢による純音聴力閾値の上昇が 2 kHz までは緩やかであるが 4 kHz および 8 kHz において加速度的に上昇する特徴が得られ, 相補的に高齢者の発話音声には 2 kHz から 4.5 kHz ではパワーが低下し, 4.5 kHz 以上の高周波帯域でのパワーが上昇する特徴が現れることが示された. 正常聴力を有する高齢者は知覚年齢が低く評価される傾向があり, 難聴者の音声の知覚年齢は, 実年齢よりも高く知覚され, 実年齢よりも知覚年齢が高い場合は嗶声がより支配的であることも明らかにされた[3]. これらのように, 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, 嗶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嗶声の重症度を表す Grade や粗造性・ガラガラ感を表す Roughness, 氣息性・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・カミを表す 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[4]. 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.

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. 本稿では, 音声の年齢知覚が聴覚の変調知覚に関係しているものと仮定して, 音声の知覚年齢と聴覚的要因 (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) の関係を調査する.

2. 年齢知覚実験

2.1. 実験刺激

実験には, 日本語小学生音声データベース (DB) (ALAGIN) および東北大 - 松下 単語音声 DB (TMW) に収録された音声データを利用した[6][7]. ALAGIN には 7 歳から 10 歳の男女の音声収録されており, 本実験では男性話者 7 名, 女性話者 10 名の計 17 名の音声を実験刺激として利用した. また, TMW には 20 代から 40 代の男女の音声収録されており, 本実験では男性話者 30 名, 女性話者 30 名の計 60 名の音声を実験刺激として利用した.

2.2. 実験方法

話者の実年齢とその話者が発する音から知覚される年齢 (知覚年齢) の関係を調べるために, 年齢知覚実験を行った. 本稿では, 二つの音声刺激セットを合わせて実験を行わず, それぞれの刺激セットで別々に年齢知覚実験を行った. ALAGIN を利用した実験では正常聴力を有する日本語母語話者 7 名 (20 代男性 5 名, 20 代女性 2 名) が参加した. TMW を利用した実験では正常聴力を有する日本語母語話者 5 名 (20 代男性 3 名, 20 代女性 1 名, 30 代女性 1 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, ヘッドホンにて音声刺激をランダム呈示し, 知覚した年齢を 1 歳刻みで回答させた.

2.3. 実験結果

図 1 に実験結果を示す. 横軸は話者の実年齢であり, 縦軸は知覚年齢を示す. 図中の赤の+は ALAGIN の結果を, 青い○は TMW の結果を示す. 赤の実線は二つの結果を合わせたものに対して得られた 2 次の回帰曲線を示し, 青の破線は ALAGIN の結果に対する回帰直線, 青の一点鎖線は TMW の結果に対する回帰直線を示す. 図 1 の結果から, 子供の音声の場合は, 実年齢

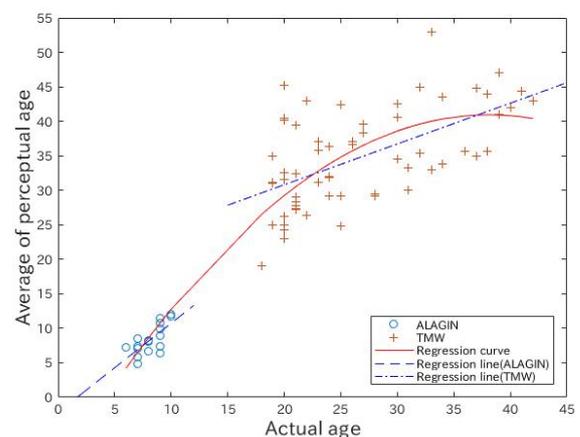


図 1 話者の実年齢と知覚年齢の関係.

と知覚年齢がほとんど近い値であることがわかる。一方、成人の音声の場合は、実年齢が上がるにつれて、知覚年齢の変動幅が広がり、飽和する傾向にあることが分かる。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[8]。図1では、これらの結果に類似した傾向がみられる。

3. 知覚年齢と音響特徴／音質評価指標の関係

3.1. 音響特徴を利用した分析

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 F_0 が特に年齢との関係があることが明らかされた。そこで、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。 F_0 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 F_0 平均を求めた[9][10]。また、スペクトル傾斜は、フレーム長10ms、フレームシフト5ms、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。図2に今回求めたスペクトル傾斜の一例を示す。図中の、青い点線は音声のパワースペクトルを示し、赤い実線は最小二乗法により求めた、スペクトル傾斜を示している。

図3に男性音声の知覚年齢とスペクトル傾斜の関係を、図4には男性音声の知覚年齢と F_0 の関係を、図5には女性音声の知覚年齢とスペクトル傾斜の関係を

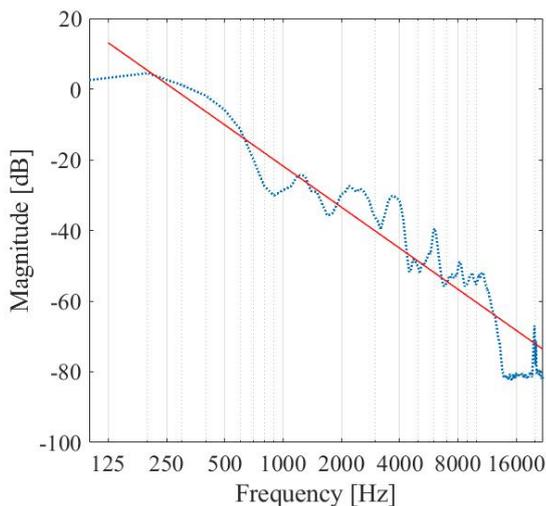


図2 スペクトル傾斜の一例。

を、図6には女性音声に対する知覚年齢と F_0 の関係を、図中の赤い+はALAGINの結果を、青い○はTMWの結果を示す。図中の赤い実線は全体の結果に対して得られた2次回帰曲線を、青の破線はALAGIN

の結果に対する回帰直線を、青の一点鎖線はTMWの結果に対する回帰直線を示す。

図3の結果より、男性音声の場合は、知覚年齢の変化とスペクトル傾斜の変化はないことが分かった。図4の結果より、男性音声の場合は、年齢が増加すると F_0 が急激に減少し、40代を過ぎると増加する傾向が見られた。図5の結果より、女性音声の場合も男性音声と同様に知覚年齢の変化に伴うスペクトル傾斜の変化はないことが分かった。図6の結果より、女性音声の場合は、年齢が増加するとともに F_0 の値も緩やかに下降していくという結果が得られた。これは、先行研究で明らかにされた話者の加齢に伴い F_0 が緩やかに低下するという結果に一致する結果となった。

3.2. 音質評価指標

まず、OROS社のSound Quality Liteを用いて、実験刺激に対する音質評価指標（ラフネス、変動強度、シャープネス）を算出した。

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60dB、搬送波周波数が1kHz、変調周波数が70Hz、変調度が1のとき、1asperと定義されており、次式のように表される[11]。

$$R = 0.3 f_{\text{mod}} \int_0^{24\text{Bark}} \Delta L(z) dz \quad [\text{asper}] \quad (1)$$

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60dB、搬送波周波数が1kHz、変調周波数が4Hz、変調度が1のとき、1vacilと定義されており、次式のように表される[11][12]。

$$F = \frac{0.008 \int_0^{24\text{Bark}} \Delta L(z) dz}{\frac{f_{\text{mod}} + 4}{4 f_{\text{mod}}}} \quad [\text{vacil}] \quad (2)$$

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60dBで、1kHzを中心周波数とする臨界帯域幅の狭帯域雑音のときに1acumと定義されており次式のように表される[11]。

$$S = 0.11 \frac{\int_0^{24\text{Bark}} N'(z) g'(z) dz}{\int_0^{24\text{Bark}} N'(z) dz} \quad [\text{acum}] \quad (3)$$

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。

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との関係を求めた。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質評価指標の平均値を用いた。

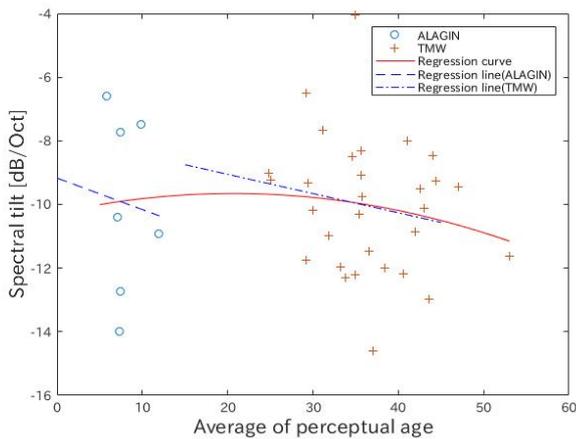


図 3 知覚年齢とスペクトル傾斜の関係 (男声).

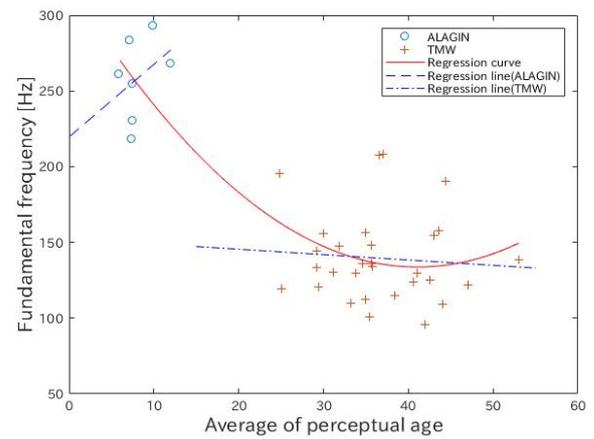


図 4 知覚年齢と F_0 の関係 (男声).

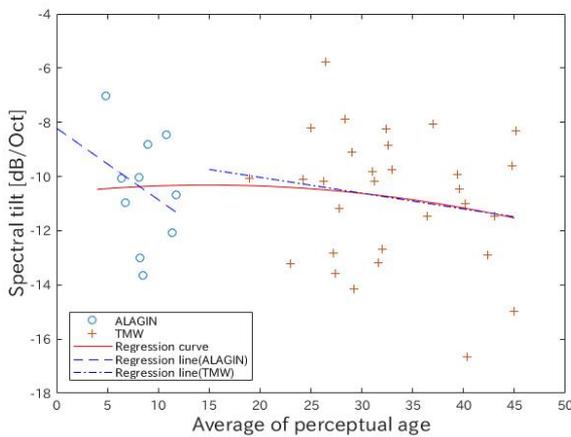


図 5 知覚年齢とスペクトル傾斜の関係 (女声).

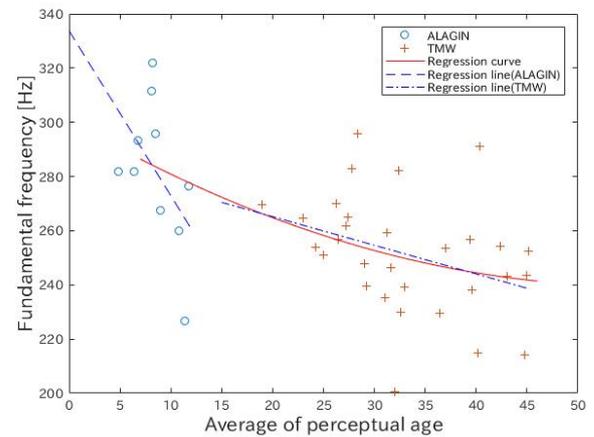


図 6 知覚年齢と F_0 の関係 (女声).

3.3. 分析結果

まず、知覚年齢とラフネスの関係を調査した。図 7 に男性音声の知覚年齢とラフネスの関係を、図 8 に女性音声の知覚年齢とラフネスの関係を示す。図中の赤の+は ALAGIN、青い○は TMW の結果を示し、赤の実線は 2 次の回帰曲線、青の破線は ALAGIN の結果に対する回帰直線、青の一点鎖線は TMW の結果に対する回帰直線を示す。これらの結果より、男性音声の場合も女性音声の場合も年齢が上がるにつれてラフネスの値が上昇していき 30 代を過ぎると降下傾向にあることがわかる。子供と成人での結果を別々に見ると、男性音声の場合も女性音声の場合も子供の頃は加齢とともにラフネスは増加傾向があるが、成人では下降傾向にあることがわかる。

次に、知覚年齢と変動強度の関係を調べた。図 9 に男性音声の知覚年齢と変動強度の関係を、図 10 に女性音声の知覚年齢と変動強度の関係を示す。図中の記号等は先の図と同様である。これらの結果より、男性音声の場合には年齢にかかわらず変動強度は変化しない傾

向が得られ、女性音声の場合は年齢が上がるにつれて変動強度が上昇する傾向が見られる。変動強度もラフネスと同様に子供と成人で別々に結果を見ると、男性音声は子供の頃は加齢とともに増加傾向を示し、成人ではわずかに下降傾向を示していることが見られる。女性音声の場合は対照的に、子供の頃は下降傾向を示し、成人ではわずかに増加傾向を示している。

最後に知覚年齢とシャープネスの関係を調べた。図 11 に男性音声の知覚年齢とシャープネスの関係を、図 12 に女性音声と知覚年齢とシャープネスの関係を示す。図中の記号等は先の図と同様である。これらの結果より、男性音声の場合は、年齢が上がるにつれてシャープネスが緩やかに下降する傾向が見えるのに対して、女性音声の場合は、男性音声の場合とは対照的に年齢が上がるにつれてシャープネスが緩やかに上昇する傾向が見られた。シャープネスも同様に別々に結果を見ると、男性音声の場合は、子供の頃は減少傾向が見られ、成人では増加傾向が見られる。女性音声の場合では、子供成人ともに減少傾向が見られる。

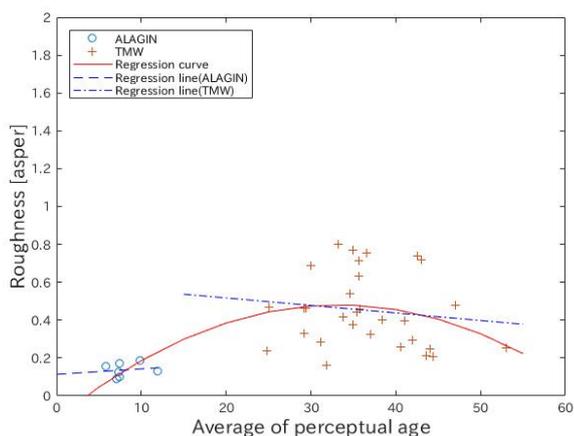


図 7 知覚年齢とラフネスの関係 (男声).

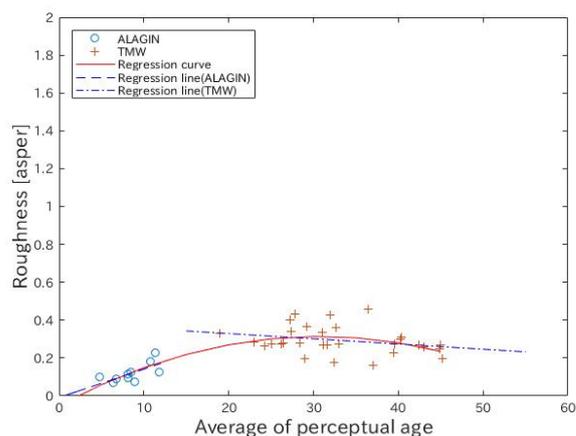


図 8 知覚年齢とラフネスの関係 (女声).

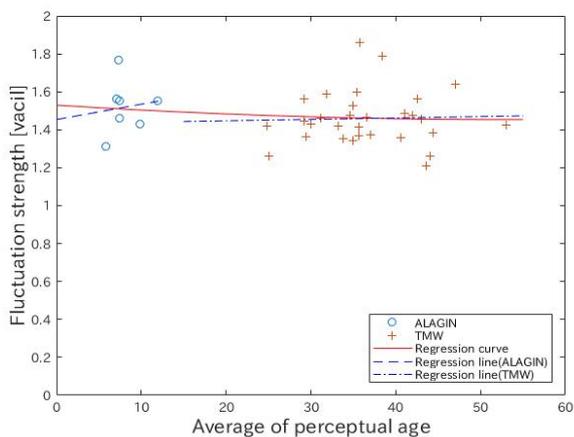


図 9 知覚年齢と変動強度の関係 (男声の場合).

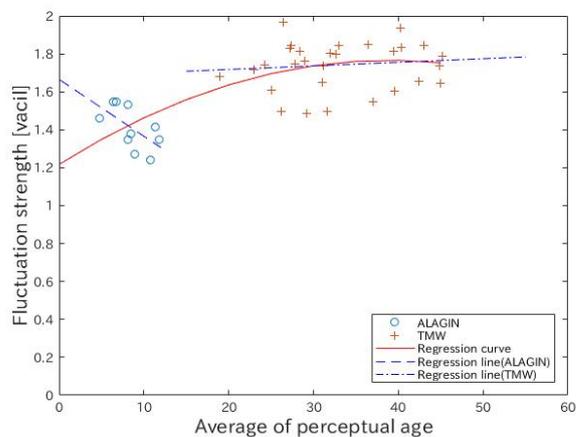


図 10 知覚年齢と変動強度の関係 (女声の場合).

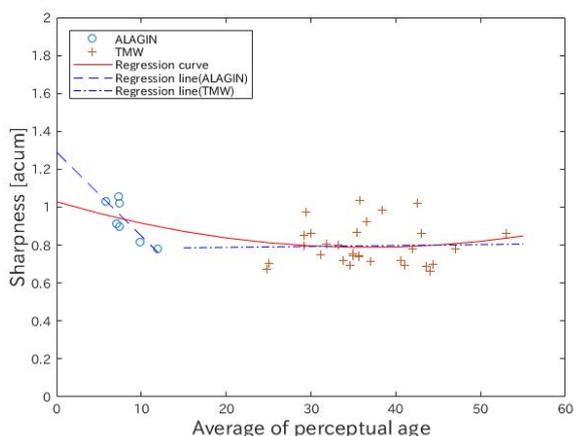


図 11 知覚年齢とシャープネスの関係 (男声).

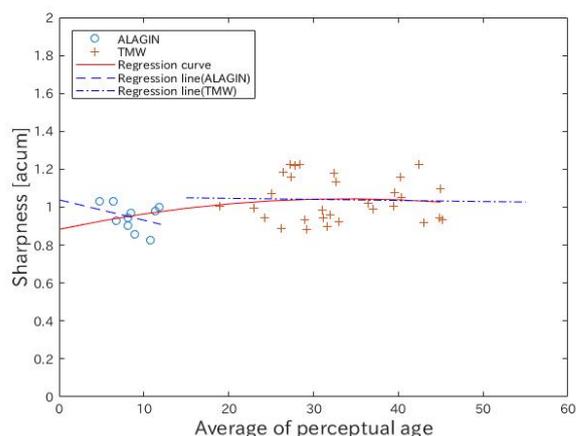


図 12 知覚年齢とシャープネスの関係 (女声).

4. 考察

知覚年齢とスペクトル傾斜, ならびに知覚年齢と F_0 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なかっ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は, 今回のデー

タを十分に説明できないといえる. そこで, 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, 男性音声の知覚年齢とラフネス, シャープネスに傾向が見られたことから, 男性音声の知覚

年齢を考える場合、ラフネスやシャープネスを用いることで説明が可能ではないかと考えられる。また、女性の知覚年齢とラフネスと変動強度、ラフネスに傾向が見られたことから、女性音声の場合は、ラフネスや変動強度、変動強度を用いることで説明が可能ではないかと考えられる。

今回の検討では、子供と成人のDBに対して別々に知覚年齢を求め、従来研究で年齢との関係が示唆された音響特徴や音質評価指標を見てきた。しかし、聴知覚のメカニズムの解明をするためには、子供と成人とを別々に知覚年齢を見るのではなく子供から高齢者までのデータを統合したうえで知覚年齢を求め、今回と同様に音質評価指標を用いて分析することで、さらに深い知見が得られるのではないかと考えられる。また、分析をする前の予想では、ラフネスは年齢とともに徐々に上昇するような傾向が得られると考えていたが、今回の検討ではラフネスが男性、女性ともに最終的に降下したような傾向が得られた。今後、前述したように高齢者のDBを含めたうえで今回のような分析をすることで、ラフネスが上昇し続けるような、傾向が得られるように考えられる。

5. まとめ

本稿では、年齢知覚に関して聴知覚メカニズムから検討するために、音声聴取による知覚年齢実験を行った。得られた実験結果をもとに、知覚年齢と音響特徴（スペクトル傾斜と基本周波数）との対応関係を調べた。さらに、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）との対応関係を調べた。その結果、実年齢と知覚年齢の対応関係から、話者の年齢が上がるにつれて年齢知覚のバラつきが見られることがわかった。このバラつきに関しては、特に成人音声の場合、35歳を起点に老若の推定結果の逆転がみられることがわかった。音声聴取による年齢知覚では、男性音声の場合、音声のラフネスや変動強度、女性音声の場合はラフネスや変動強度、シャープネスが知覚年齢に関係していることがわかった。今後は、子供から高齢者までの音声データを集めたうえで、本検討と同様に分析していくことで、声の年齢知覚のメカニズムを解明する予定である。

謝辞

本研究は科研費・新学術領域（No.18H05004）の支援を受けて行われた。

文 献

- [1] 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—感情・パラ言語情報・個人性の科学—, コロナ社, 2014.

- [2] 粕谷英樹, “性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.
- [3] 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.
- [4] 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.
- [5] 日本音声言語医学会, 新編声の検査法, 医歯薬出版会社, 2009.
- [6] 東北大 - 松下単語音声データベース(TMW)
<http://research.nii.ac.jp/src/TMW.html>
- [7] CIAIR 子供の声データベース(CIAIR-VCV)
<http://research.nii.ac.jp/src/CIAIR-VCV.html>
- [8] 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.
- [9] 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.
- [10] Kawahara, H., Masuda-Katsuse, I., and de Cheveigne, A., Restructuring speech representations using a pitch adaptive time-frequency smoothing and an instantaneous-frequency based on F0 extraction: Possible role of a repetitive structure in sounds, Speech Communication, Vol. 27, pp. 187-207, 1999.
- [11] Fastl, H., Zwicker E., “Psychoacoustics : Facts and models, third edition” Springer, pp. 257-264, 2006.
- [12] Fastl H., “Fluctuation strength and temporal masking patterns of amplitude modulated broadband noise,” Hearing Research, 8, pp. 59-69, 1982.