

日本心理学会 第85回大会 オンライン開催

ささやくシルエツト： 声の音色の純粹培養

As we speak: pure culture of the voice timbre

大学入試センター 研究開発部 内田照久

動画版には 音声Demo が多く収録されています
下記の Drop Box からダウンロードしていただいて
どうぞそちらをご覧ください (2021.12.22 追記)

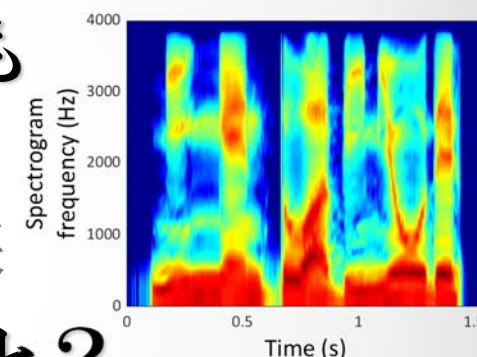
https://www.dropbox.com/s/sxs8qxtw7of6moc/Uchida_JPA2021.zip?dl=0

2021/9/1(水)～8(水)

1. はじめに “声” をより豊かに！ さらに自由に！

1. 音声のアクセントやイントネーションを担う基本周波数や調波構造がなくても人は何を言っているか 聴き取れる

声の音色にあたるスペクトル包絡に情報



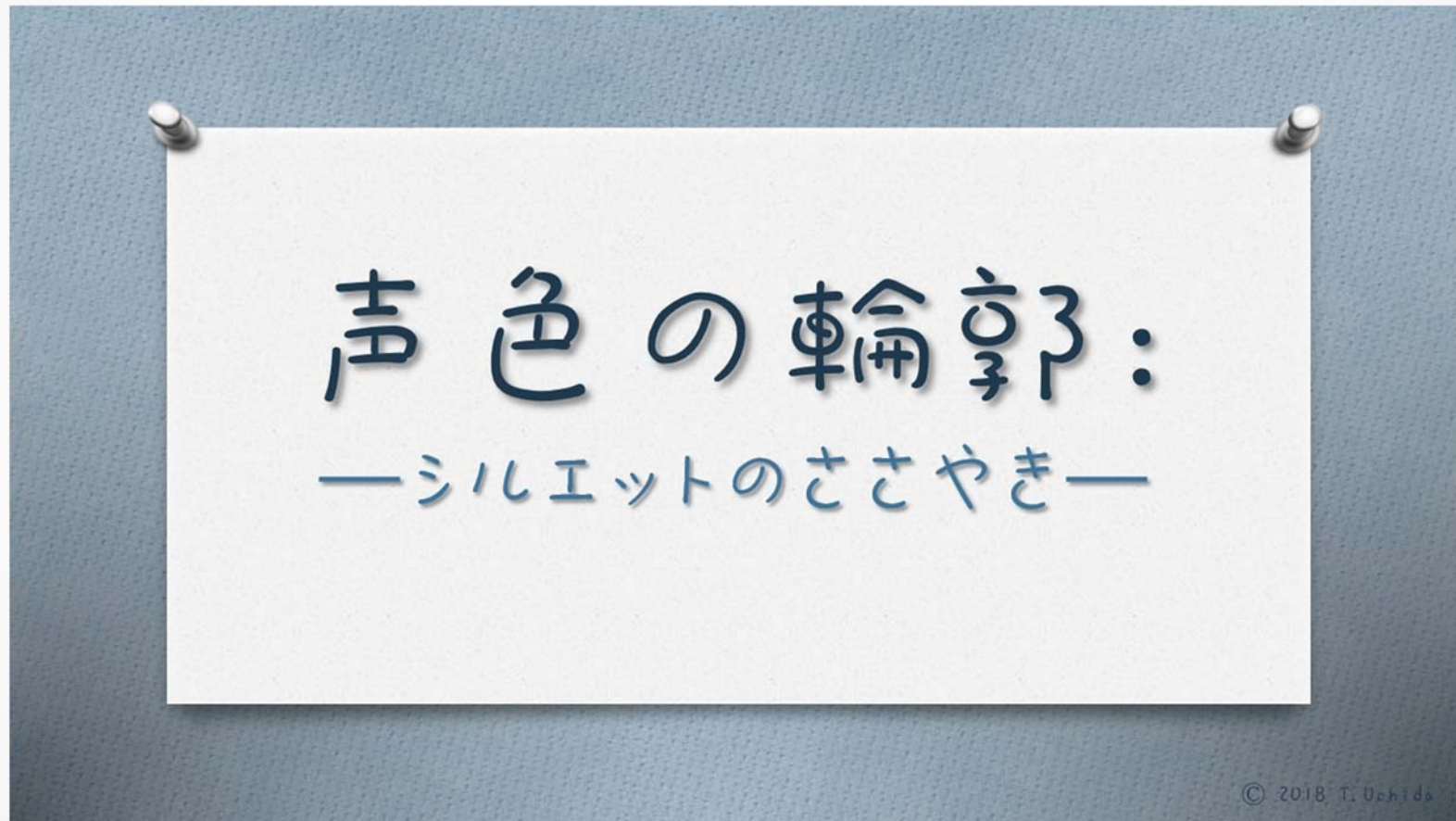
2. 声の音色の形を 白色雑音に刻み込むと？

「声色の輪郭: —シルエットのささやき—」

3. ささやき声の生成法を提案 Phantom Silhouette方式
高品質vocoder: WORLD*をベース ⇒ 特殊効果

* M. Morise. (2016). *Speech Communication*, 84, 57-65.

2. 声の“音色”を 純粹培養したら？



3. 実用的なささやき声の生成法: — Phantom Silhouette方式の提案 —

1. 通常発話の音声を“ささやき声”に
任意の箇所を後から加工も
2. 実用的でパラメトリックな簡便法を提案
高品質vocoder: WORLDを利用
3. 主観評価実験と提案手法の改良
対比較による聴覚的な評価実験

内田照久・森勢将雅 (2019). 日本音響学会 2019年 秋季研究発表会 講演論文集, 973-976.

内田照久・森勢将雅 (2020). 日本音響学会 2020年 秋季研究発表会 講演論文集, 771-774.

3.1 疑似ささやき声の生成

1. 通常音声のスペクトル包絡を抽出
2. ささやき声に聞こえるように特徴を操作☆
3. 声帯音源信号の代わりに白色雑音で駆動

☆スペクトル包絡の変換操作

- (a) F_1, F_2 の上方シフト*: スペクトル周波数軸伸縮
- (b) 氣息音成分の補填: 高域スペクトルに重みづけ
- (c) スペクトルの低域抑圧*: 音韻了解性に配慮**

* 松田, 森, 粕谷, 音響学会誌, 56 (7), 477-487, 2000.

** Kishida, et al., *proc. of ICP 2016*, 303, 2016.

Phantom Silhouette 方式

ハロウィーンのお化けのイメージ

通常音声のスペクトル包絡 $\Rightarrow$ 雑音駆動による無声化

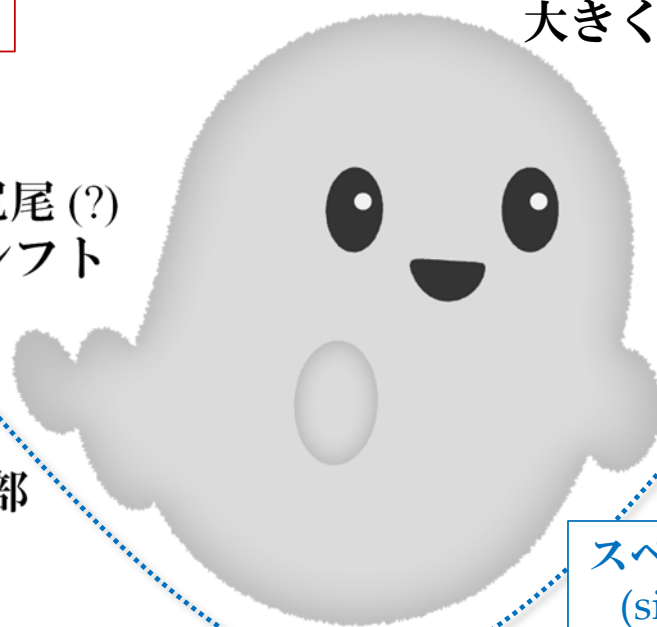
音声のファントム化
(phantomization)

ちょっぴり
跳ね上がった尻尾 (?)
(a) $F1, F2$ 上方シフト

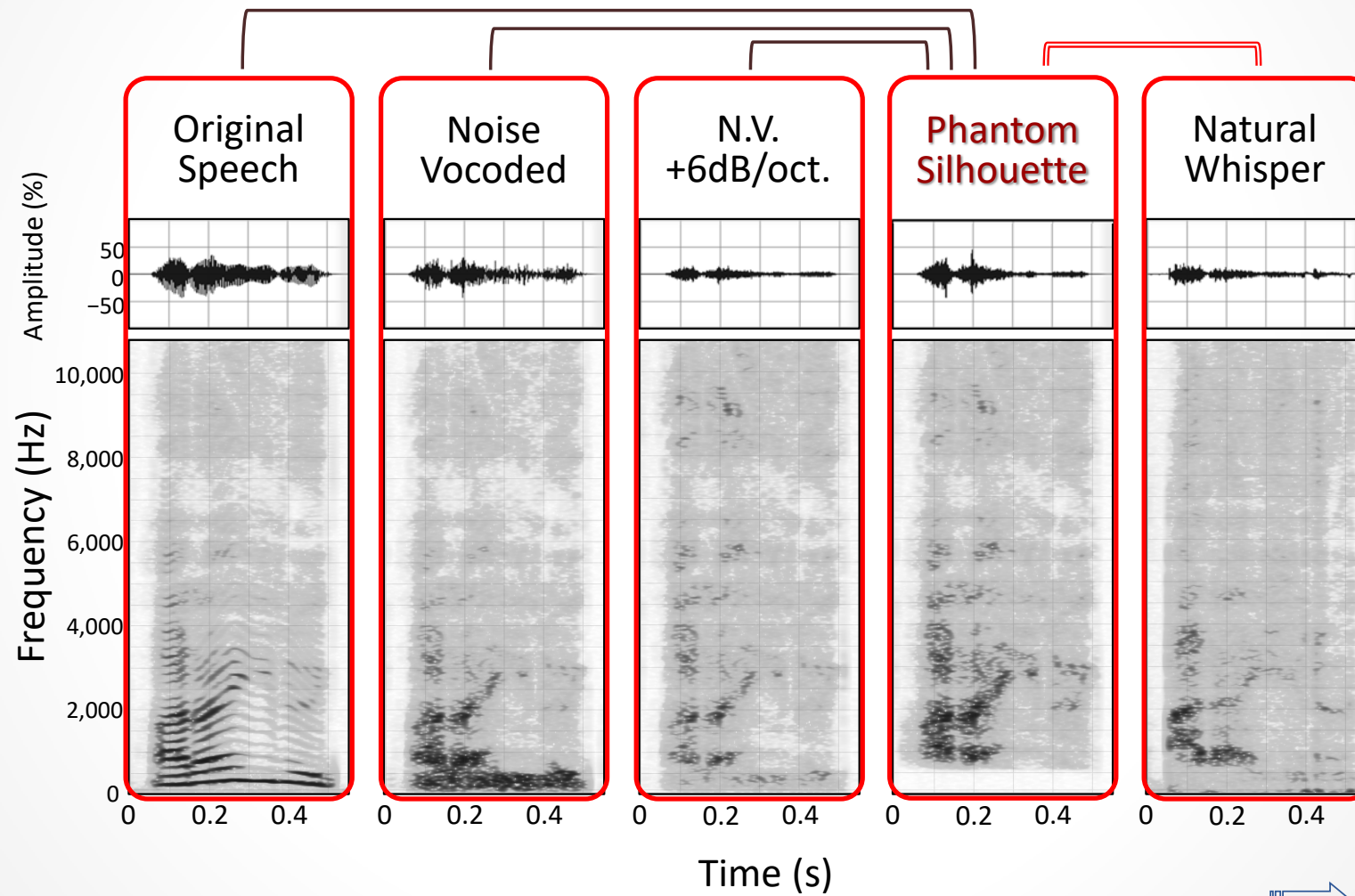
細く消えていく脚部
(c) 低域抑圧

大きくデフォルメされた頭部
(b) 高域補填

スペクトルのシルエット補正
(silhouette compensation)



対比較による主観評価実験へ



3.2 疑似ささやき声の主観評価

1. 原音声: ATR音素バランス文より 8文章
ナレーター事務所に所属: 男声/女声 各4名
“通常発話”と“ささやき声”を収録

(男声4名+女声4名)×2 スクリプト×(通常発話+ささやき声)=計 32 音声

2. 比較条件: (P.S.: Phantom Silhouette方式)

- | | |
|-----------------------|------------------|
| (1) 原音声 (通常発話) | — P.S. (通常発話ベース) |
| (2) 雑音駆動音声 | — P.S. (通常発話ベース) |
| (3) 雑音駆動音声 (+6dB/oct) | — P.S. (通常発話ベース) |
| (4) 実際のささやき声 | — P.S. (通常発話ベース) |

(ラウドネス・レベル; 原音声: -24LKFS, ささやき系: -28LKFS)

3. 参加者: 国公立大学1年生100名 (男: 80, 女: 20)

8群に分割・全64音声対も8分割

群別に音声対を配置したCD-Rを作成

4. 評価項目:

ささやき声の評価を含む声質表現語,
声の自然性 など に関する計20項目

5. 手続き:

携帯型CDプレーヤからイヤホンで音声提示

比較対の2つの音声を続けて一回提示

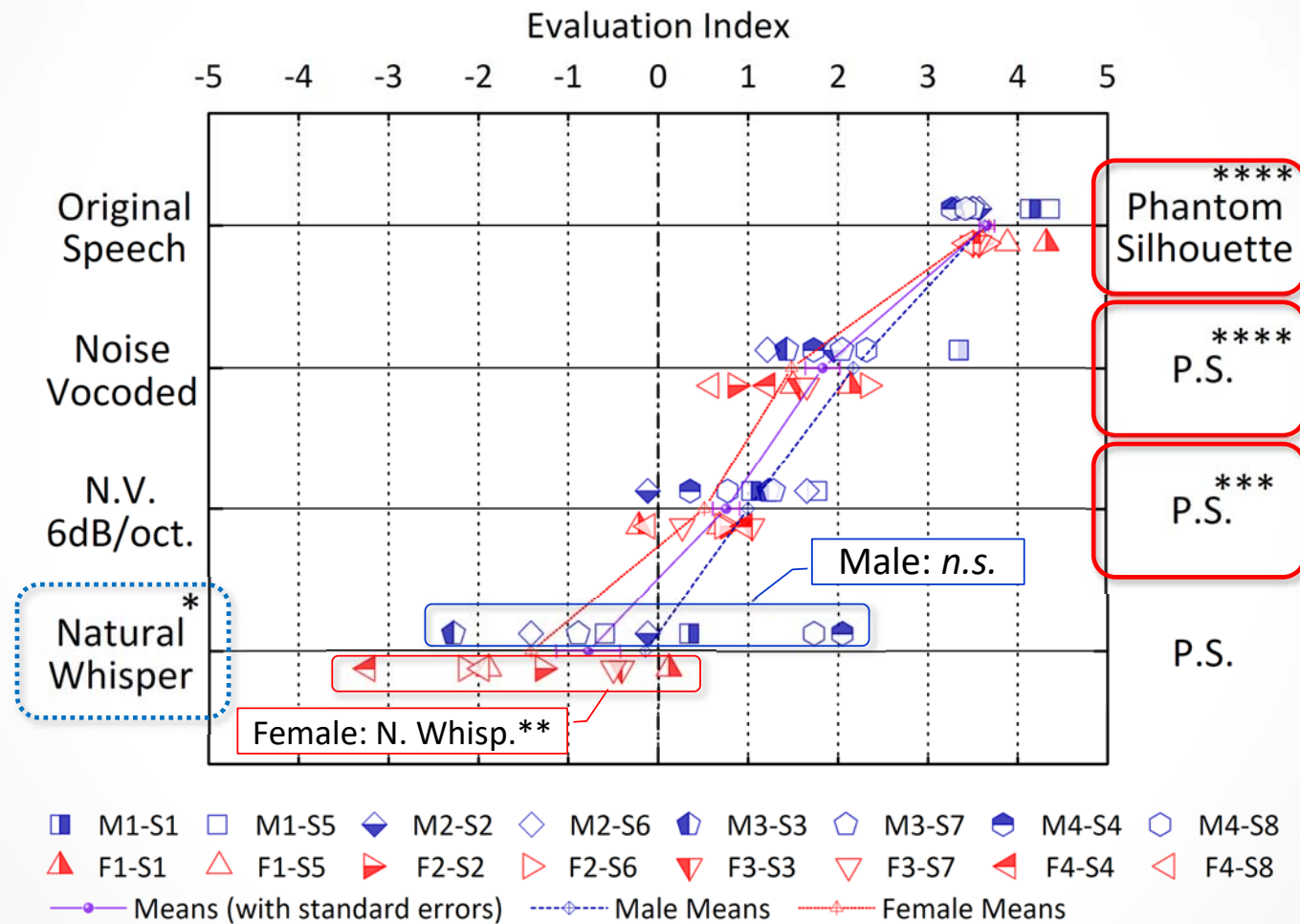
各項目に どちらがあてはまるか10段階で評価

対ごとの20項目の評価時間は60秒

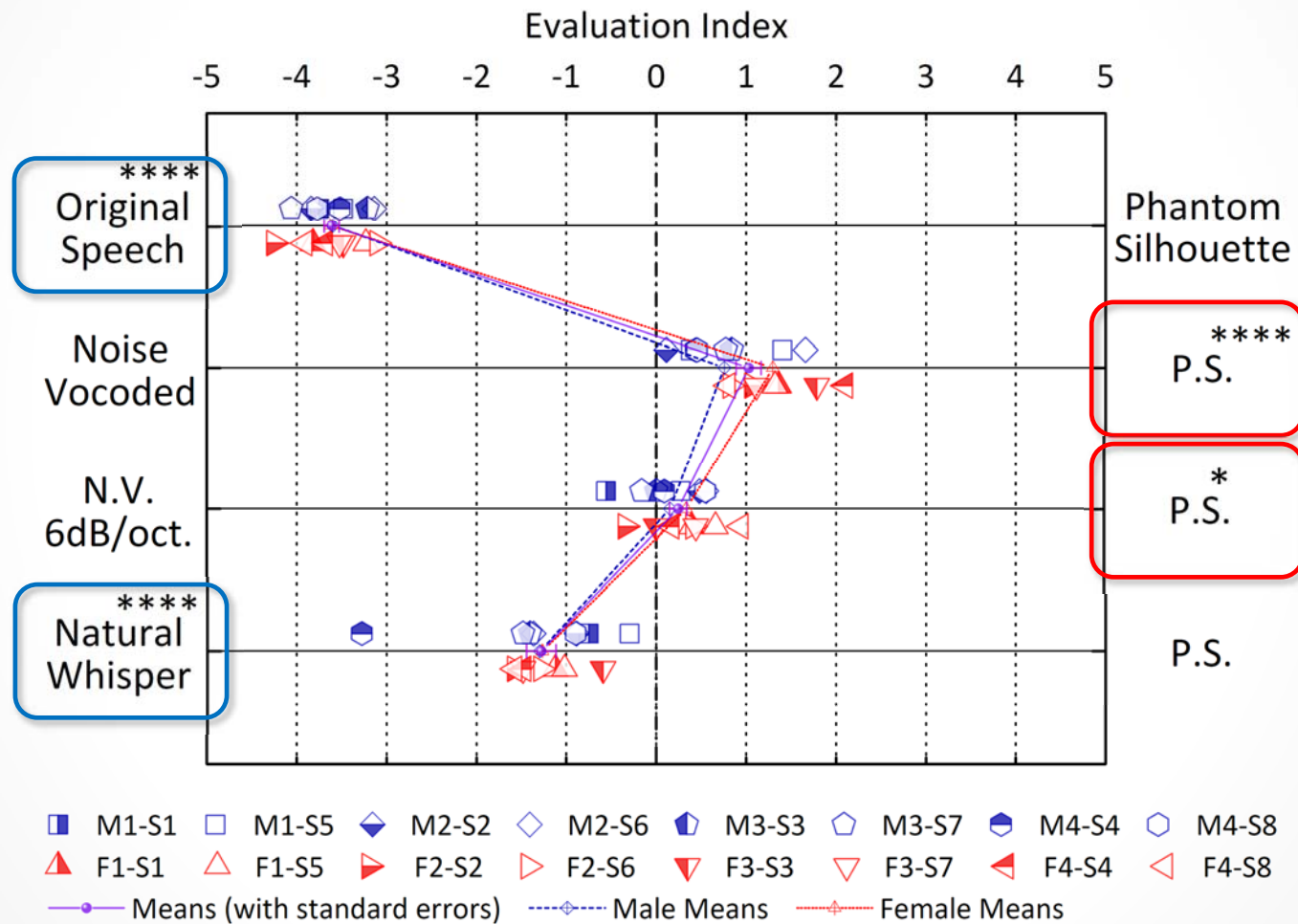
【話し方・話し手の比較 (60秒)】 [この冊子に直接マーク(●)して下さい]

	【No.1】にあてはまる ←-----					-----→	【No.2】にあてはまる				
【評定練習】	非常に ①番にあてはまる	たいへん ①番にあてはまる	かなり ①番にあてはまる	比較的 ①番にあてはまる	どちらか ①番にあてはまる といえ	どちらか ②番にあてはまる といえ	比較的 ②番にあてはまる	かなり ②番にあてはまる	たいへん ②番にあてはまる	非常に ②番にあてはまる	
高い声	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
男性的な声	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
かすれた声	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
落ち着いた声	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
迫力のある声	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
若い感じの声	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
太い声	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
張りのある声	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
鼻声	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
うら声	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
ささやき声	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
早口な	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
自然な	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
わかりやすい	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
はっきりした	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
聞き取りやすい	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
体の大きい	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
背の高い	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
太った	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
がっしりした	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	

Whisper: ささやき声



Naturalness: 自然性



3.3 提案法の改良: Phantom Silhouette 2

(a) F_1, F_2 帯域の上方シフト量の改良

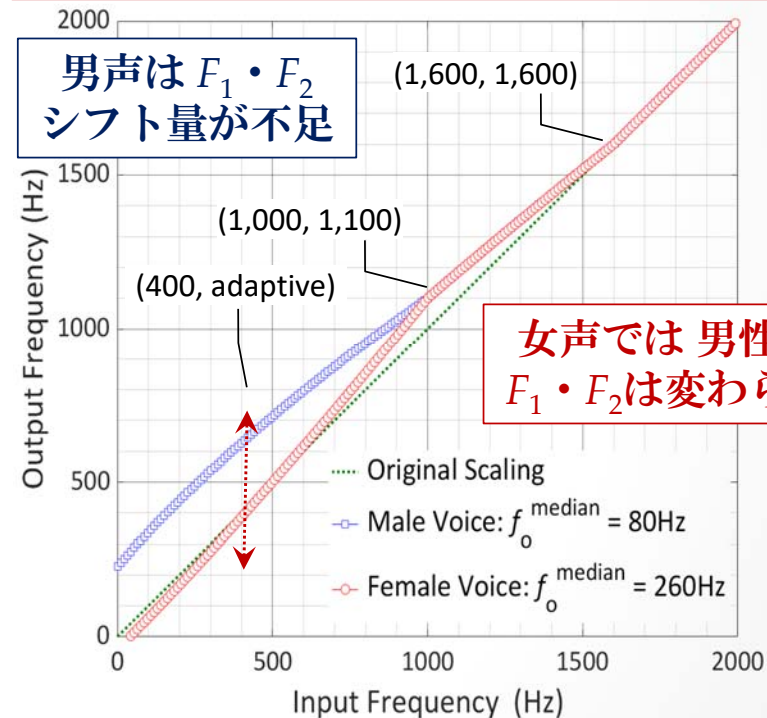
スペクトル周波数軸の非線形区分伸縮 (on ERB Scale)

Table 1 通常母音とささやき母音のフォルマント周波数と比

		/i/	/e/	/a/	/o/	/u/
Modal	F_1	302	541	654	458	375
	F_2	2,057	1,784	1,060	807	1,208
	F_3	3,004	2,452	2,375	2,379	2,165
	F_4	3,486	3,341	3,293	3,066	3,129
	F_5	4,526	4,589	4,363	4,064	4,146
Whispered	F_1	336	695	813	559	617
	F_2	1,957	1,763	1,227	906	1,303
	F_3	2,954	2,315	2,281	2,487	2,281
	F_4	4,299	3,199	3,220	3,162	3,151
	F_5	5,342	4,606	4,510	4,398	4,466
Ratio	F_1	1.11	1.28	1.24	1.22	1.65
	F_2	0.95	0.99	1.16	1.12	1.08
	F_3	0.98	0.94	0.96	1.05	1.05
	F_4	1.23	0.96	0.98	1.03	1.01
	F_5	1.18	1.00	1.03	1.08	1.08

発声者 A の ARX 分析による 1s の音声の平均値, 上段から通常母音, ささやき母音, ささやき母音/通常母音の比を示す。

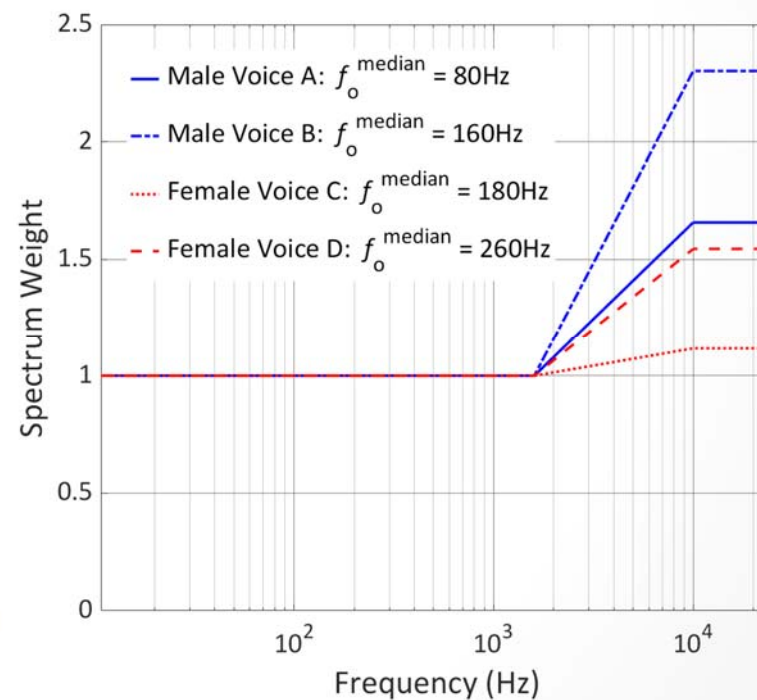
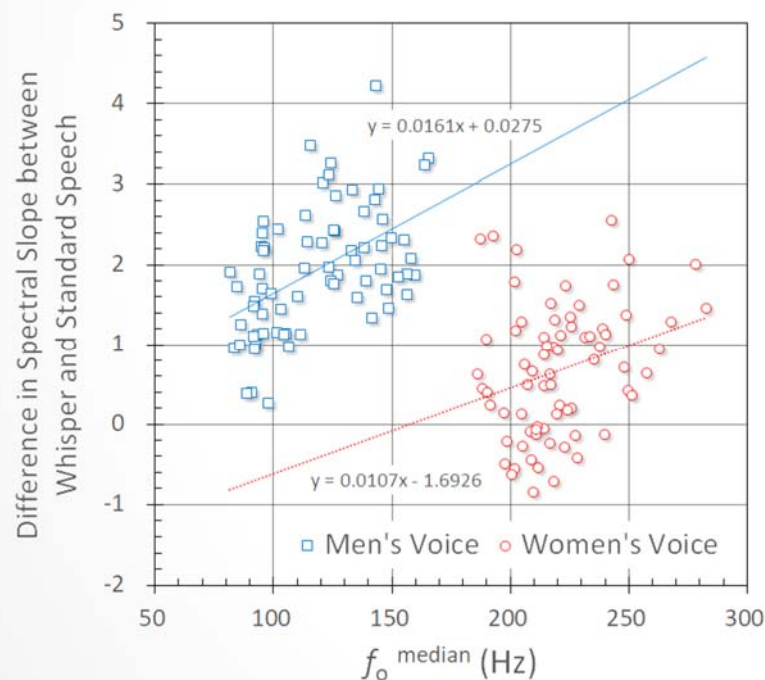
f_0 median に応じて $F_1 \cdot F_2$ シフト量を制御



松田, 森, 粕谷, ささやき母音のフォルマント構造,
音響学会誌, 56, 477-487, 2000.

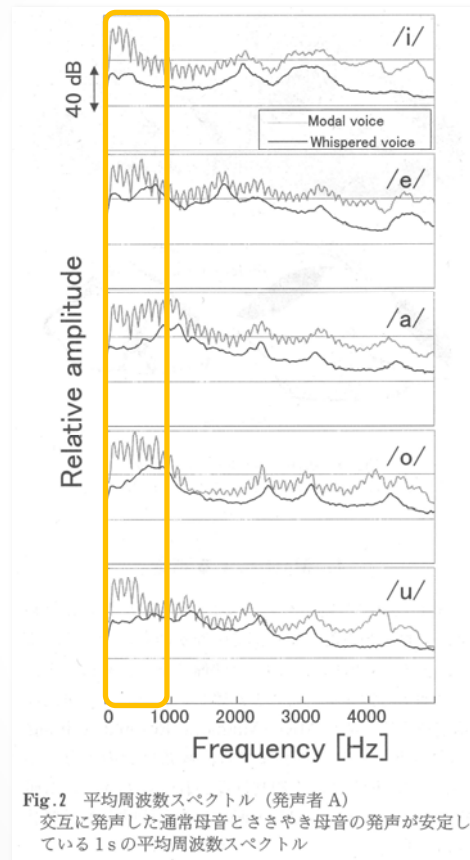
(b) 高域の気息音成分の補填量の適応的補正

1. 男声で 高域補填が必要
2. 男女とも f_0 が高い程 要:補填
 $\Rightarrow f_0$ median で補填量を補正



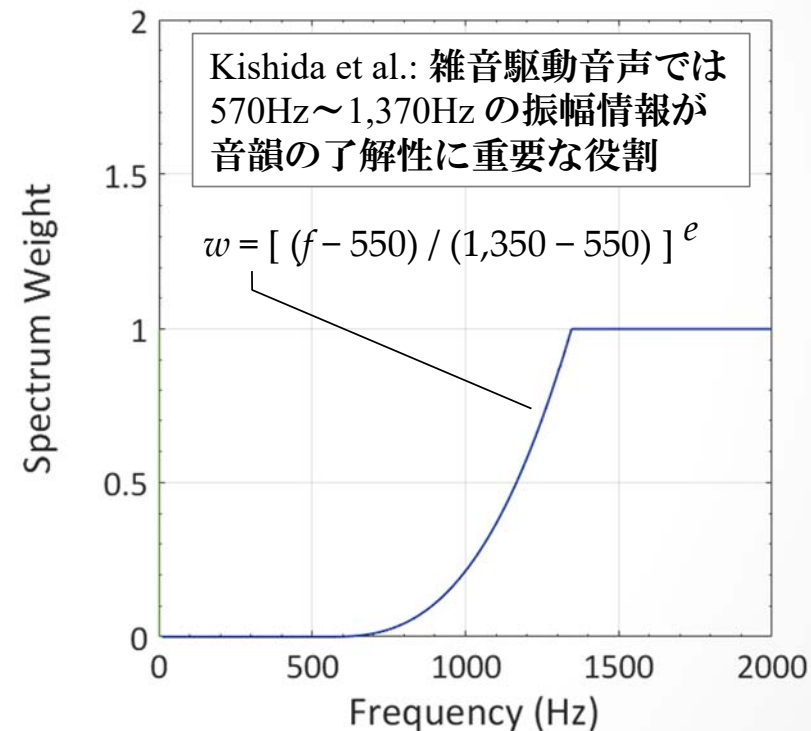
(c) 低域スペクトルの抑圧対象の拡大

スペクトルへの重みづけにより合成時に抑圧を実現



Truncated Noise-Vocoded Speech

内田, 音講論集(秋), 353-354, 2017.



松田, 森, 粕谷, ささやき母音のフォルマント構造,
音響学会誌, 56, 477-487, 2000.

Uchida, T. & Morise, M. (2021). A practical method of generating whisper voice: Development of phantom silhouette method and its improvement. *Acoustical Science & Technology*, **42** (4) 214-217.

3.4 まとめ

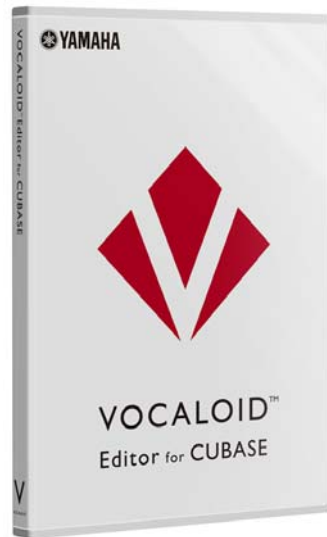
1. 変換パラメタの修正・適応化により改善
聴感上の“ささやき声らしさ”を具現化
自然性には改善の余地
2. 駆動音源の雑音種別の影響も少ない
White Noise $\Leftrightarrow$ Modified Velvet Noise *
3. 音声の生成面からの改良・拡張は未達成
ささやき時の下咽頭腔の共鳴特性**の反映

* M. Morise, “Modification of velvet noise for speech waveform generation by using vocoder-based speech synthesizer,” *IEICE transactions on information and systems*, **E102-D** (3), 663-665, 2019.

** K. Honda et al., “Visualisation of hypopharyngeal cavities and vocal-tract acoustic modelling,” *Computer Methods in Biomechanics and Biomedical Engineering*, **13** (4), 443-453, 2010.

内田照久・森勢将雅・中島 祥好 (2017). 声を操る！－STRAIGHT-WORLDにようこそー
[チュートリアル・ワークショップ] 日本心理学会 第81回大会発表論文集, TWS(3).

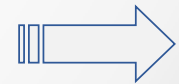
4. 発展 4.1 Vocaloid の歌声も ささやき声化!



Quasi-Whispered



邪道? 特殊効果?



4.2 朗読の声も ささやき声化!

“声”をより豊かに! さらに自由に!
どうもありがとうございました

新美 南吉:「ごんぎつね」より*



ささやき声化



原音声



* 高道 慎之介, 中田 亘, 郡山 知樹, 丹治 尚子, 井島 勇祐, 増村 亮, 猿渡 洋,
“J-KAC: 日本語オーディオブック・紙芝居朗読音声コーパス,”
情報処理学会研究報告, Vol.2021-SLP-137, No.14, 2021.