

研究背景

- ビデオ会議の普及 → QoEの重要性 → 音声品質評価も重要
- 客観品質評価技術：PESQ/UTMOS などは有用だが限界あり[1]
- 主観評価：MOSは簡便・標準化されているが微細な差を捉えにくい[2]
- 書き起こし (dictation) による評価：
 - CER/WER が聞き取りやすさを反映
 - 従来は音声の1回提示が主流[3]
- しかし実際のビデオ会議では複数回聞いて書き起こし(ex. 議事録作成)

→ 誤り率だけでなく聞き直し回数・回答時間も重要

研究命題

パケット損失による音質劣化は

- 日本語音声の聞き取り精度 (CER: Character Error Rate)
- 聞き取り負荷 (聞き直し回数・回答時間)

にどのような影響を与えるか。

聞き取り精度 CER_{first}：初回聴取時のCER

CER_{last}：最終聴取時のCER

聞き取り負荷 FLC (Final Listen Count)：最終的な聞き直し回数

T_{ans} (Answer Time)：解答時間

実験手法

参加者 日本語話者20名
男性12 女性8 平均年齢26.1歳

課題 劣化音声をヘッドフォンで聴取→ひらがなで書き起こし
最大6回まで再生可
各再生ごとに回答テキストと回答時間を記録
回答例：Table1

刺激 データベース：ASJ-JIPDEC (VOL1) [4]
文章数：48 (男声24・女声24)
モーラ数：26-35、発話速度：7.0 - 8.0 mora/s
Packet Loss Ratio
PLR (パケット損失率)：0, 10, 20, 40%

装置 MacBook Air + MAX 8.6.2 GUI
Focusrite Scarlett Solo + Sony MDR-7506

Table1: 解答例		
	テキスト	CER
元の文章	でしにうでをささえられながらさいごまでひきつづけた	0.00
1回目	でしにうでを せんまで ひきつづけた	0.56
2回目	でしにうでをさえられながら せんまでひきつづけた	0.28
3回目	でしにうでをさえられながら さいごまでひきつづけた	0.16
4～6回目	(回答なし)	

結果

1. PLRの影響

PLRの増加に伴い

CER_{first} ↑ FLC ↑
CER_{last} ↑ T_{ans} ↑

CER_{last} は CER_{first} に比べ
PLRの影響をより明確に反映

統計検定

全指標PLRの主効果あり*
*Friedman検定

PLR40% は他条件より有意に高い*
*Wilcoxonの符号順位検定 (Holm 補正)

2. 聞き直し回数とCER

聞き直し回数 ↑ → CER ↓

1→3回目で改善大

4回目以降は改善小

統計結果

FLC 2-4：
すべての比較で有意差* (p < .05)

FLC 5-6：
1-2, 2-3, 3-4回で有意 (p < .01)

4-5, 5-6回では有意差なし

*Wilcoxonの符号順位検定 (Holm 補正)

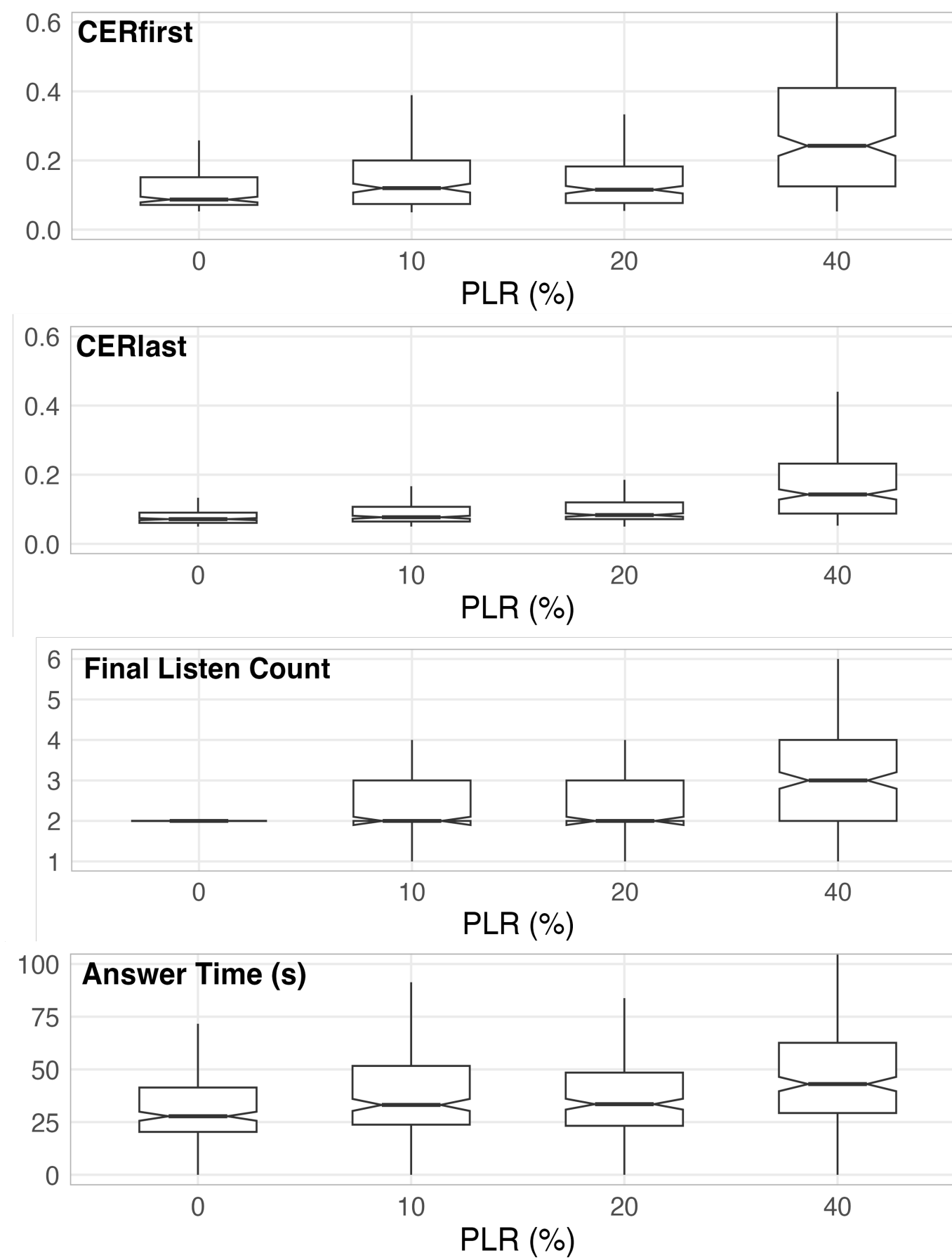


Fig.1 各 PLR 条件における 4 指標の中央値および分布の広がりを示す箱ひげ図。ノッチは中央値の95%信頼区間を表す。

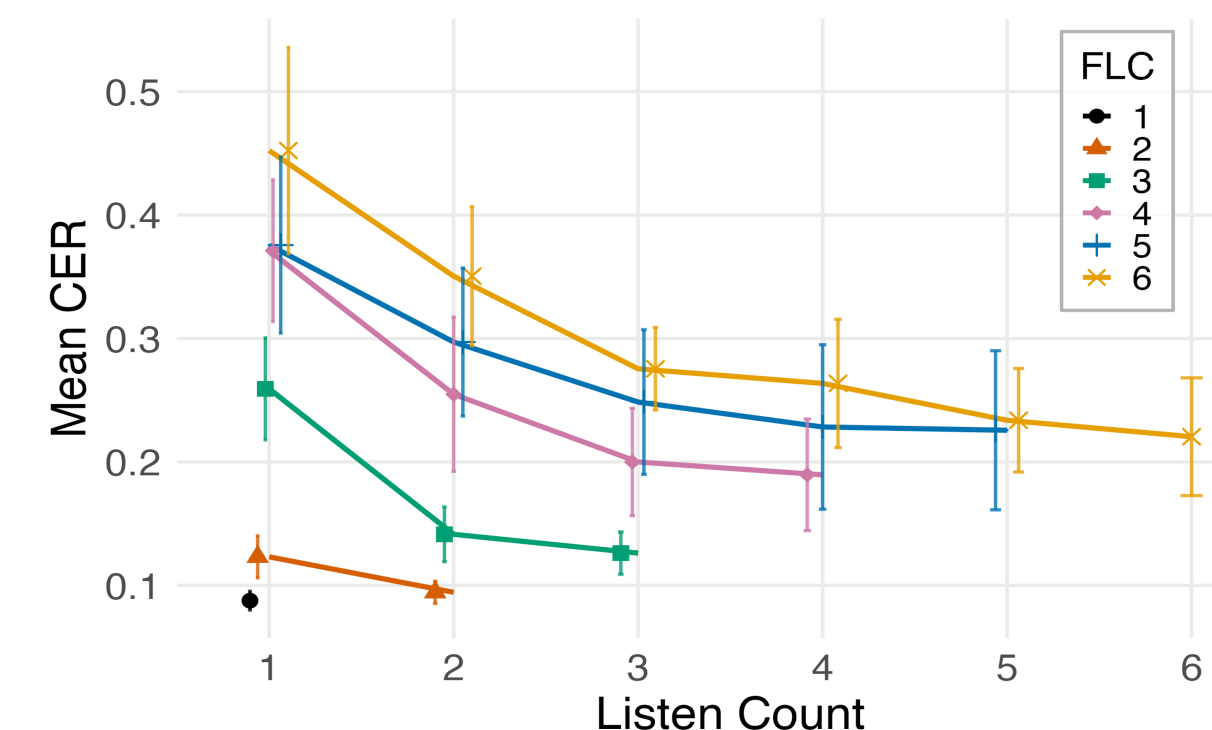


Fig.2 聞き直し回数に伴う平均 CER の推移。
誤差範囲は 95%信頼区間を表す。

聞き直し回数4回の場合のPLRの影響

1-2 回目：
PLR 0%, 10%, 20%
の差は明確でない

3回目以降：
PLR 20% > 10% > 0%
の傾向がより明瞭

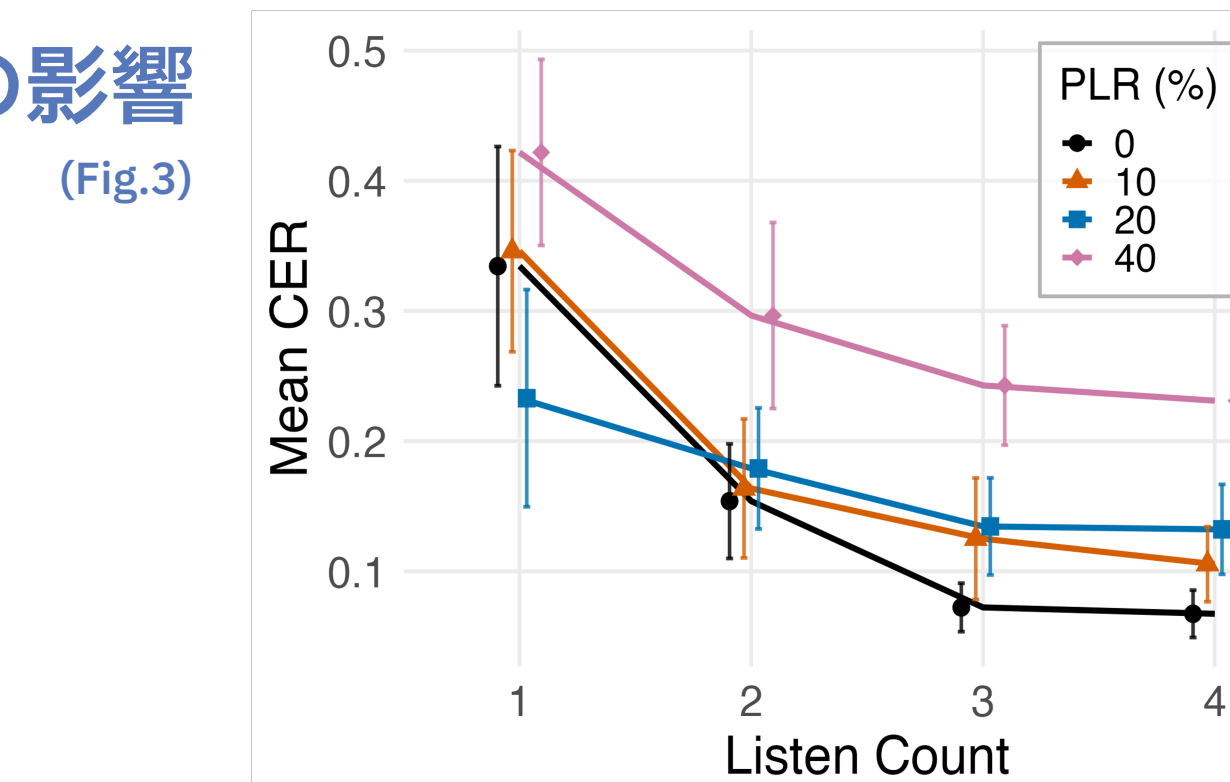


Fig.3 聞き直し回数が 4 回の条件における PLR 別の平均 CER の推移。誤差範囲は 95%信頼区間を表す。

Table2: 順序ロジスティック回帰モデルの AIC 比較

	AIC	Δ AIC
CER _{last} + FLC + T _{ans}	2415.32	0.00
CER _{last} + FLC	2415.61	0.29
CER _{last}	2440.66	25.34
CER _{first} + FLC + T _{ans}	2521.38	106.06
CER _{first} + FLC	2524.07	108.75
CER _{first}	2543.18	127.86

考察と結論

初回聴取時の精度よりも、複数回の聞き直し後の最終精度 (CER_{last}) の方が、PLR による音声品質差をより明瞭に反映することが示された。特に CER_{last} では低い PLR 条件間でも差が検出され、音声劣化の影響をより安定して捉えられる可能性が示唆された。

聞き取り精度は主に 3 ~ 4 回目までの聞き直しで改善し、それ以降では改善が小さくなる傾向が確認された。これは繰り返し聴取によって音響手掛かりが統合され、達成可能な理解水準に近づく過程を反映していると考えられる。

CER_{last} (結果) と FLC (聞き直し回数) を併用することで、PLR 条件の違いをより適切に特徴づけられることが示された。CER_{last} が理解の結果を示す指標であるのに対し、FLC は理解に至るまでの聴取負荷 (過程) を反映する指標として機能すると考えられる。

今後の課題

他の劣化要因との比較 / ASRとの対応関係の検討 / 言語的・音声的要因の影響 / 個人差の分析

参考文献

- [1] Torcoli et al., IEEE/ACM Transactions on Audio Speech, and Language Processing, vol. 29, 1530–1541, 2021.
- [2] 岸田, 日本音響学会誌, vol. 80, no. 7, 417–422, 2024.
- [3] Kang et al., Language Learning, vol. 68, no. 1, 115–146, 2017.
- [4] 日本音響学会研究用連続音声データベース (ASJ-JIPDEC), 2007.