



第102回SLUD研究会

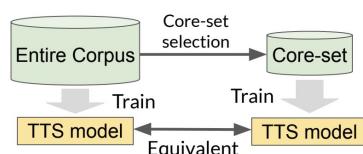
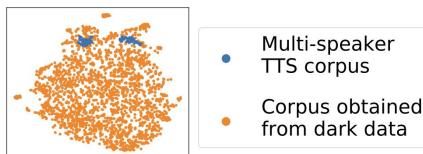
INTERSPEECH 2024 參加報告

東京大学 猿渡・齋藤研究室 博士1年 関健太郎

自己紹介

- 名前：関 健太郎
- 所属：東京大学 猿渡・齋藤研究室 博士1年生
- 研究分野：音声合成・音声強調

研究テーマ1：インターネットデータからの音声合成

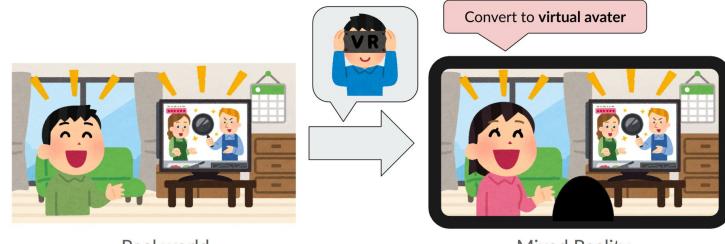


インターネット上の
幅広いデータを
[Seki+2023]

うまく間引きして
効率的に利用
[Seki+2024]

INTERSPEECH2024にて発表

研究テーマ2：MRのためのボイスチェンジャー



没入感を保つための信号処理技術 [Seki+2024]

INTERSPEECHとは？

- 音声分野では ICASSP と並ぶ二大トップ国際会議
- 対象分野：音声に関する全般
 - 音声認識、音声合成、対話、音源分離・強調、音声科学 etc...



Opening Ceremony より引用

開催地：Kos, Greek

一言でいえば「リゾート地」

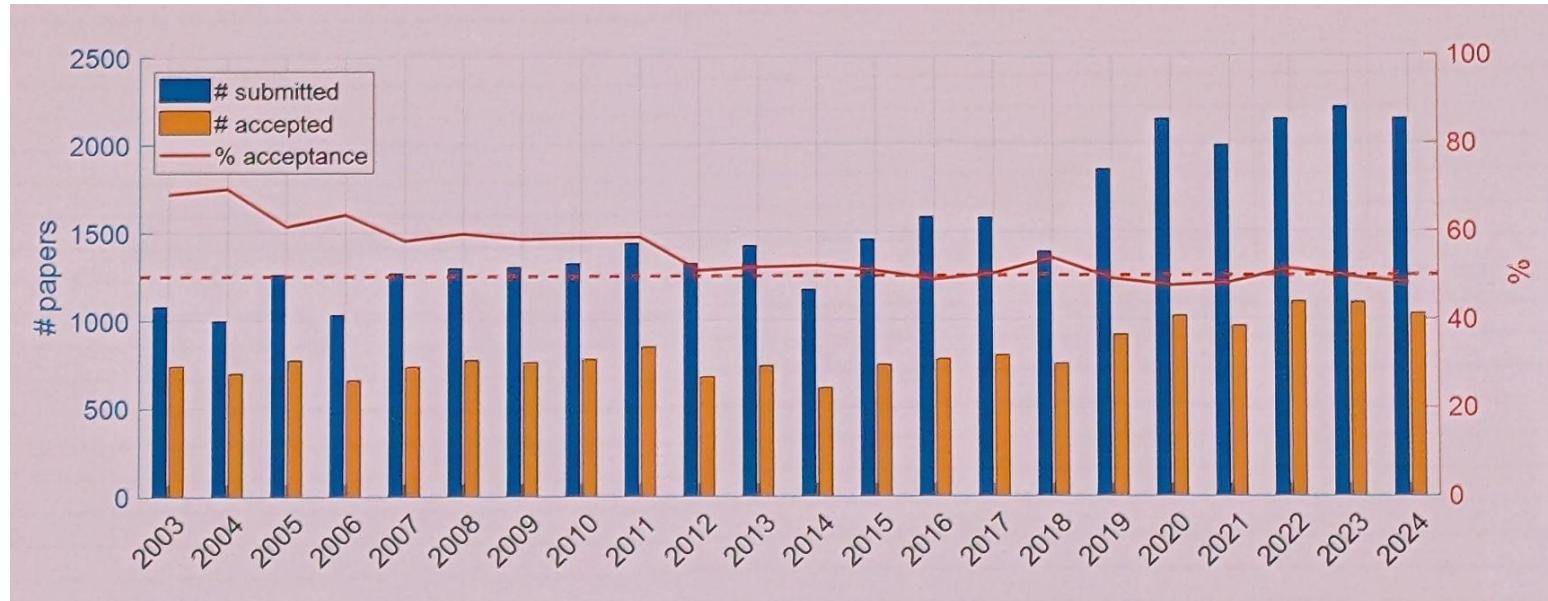


数字で見るINTERSPEECH 2024

- 採択率：48.22%（採択1031件／提出2138件）
 - desk rejectは除いた数字
 - （参考）ICASSP 2024 : 2812/5796
- ページ数：5ページ (Short Paper)
 - ※最後1ページはAcknowledgementとReferenceのみ
- 査読プロセス
 - 1780人の査読者が6544件の査読を実施
 - 各論文に3件以上の査読（※6本の例外では2件の査読）
- 参加者：1800人弱
- セッション数：106 technical sessions + 11 special sessions
 - 42 Show and Tell in 4 sessions

採択性の遷移

- 直近10年は50%前後で推移、ここ数年のsubmission数は安定している傾向



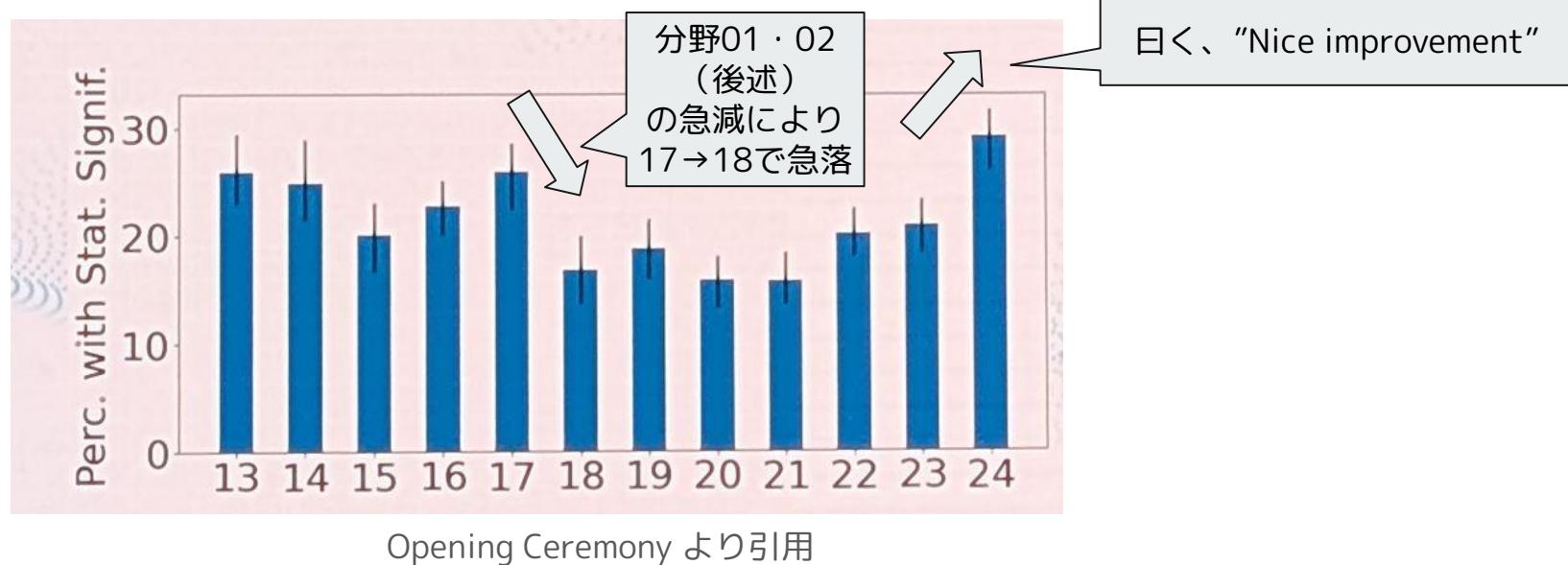
Opening Ceremony より引用

今回導入された試み

- Double Blind Review の導入
 - 査読者からはpaperの著者が見えない
- (上に付隨して) Anonymous Periodの導入
 - 締め切り1ヶ月前～採否通知の期間、非匿名版の公開禁止 (arXivもダメ)
 - 個人的な所感：Preprint文化との相性の悪さ（特に機械学習系の論文）
 - 1ヶ月早く執筆し終わった著者によるarXiv投稿も見受けられた
 - 来年については不明
 - Paper Submissionのページには見受けられず
- Blue Sky Trackの導入
 - 萌芽的研究を扱うトラック（実験結果が不足していても受け入れる）
 - 13本のうち2本が採択

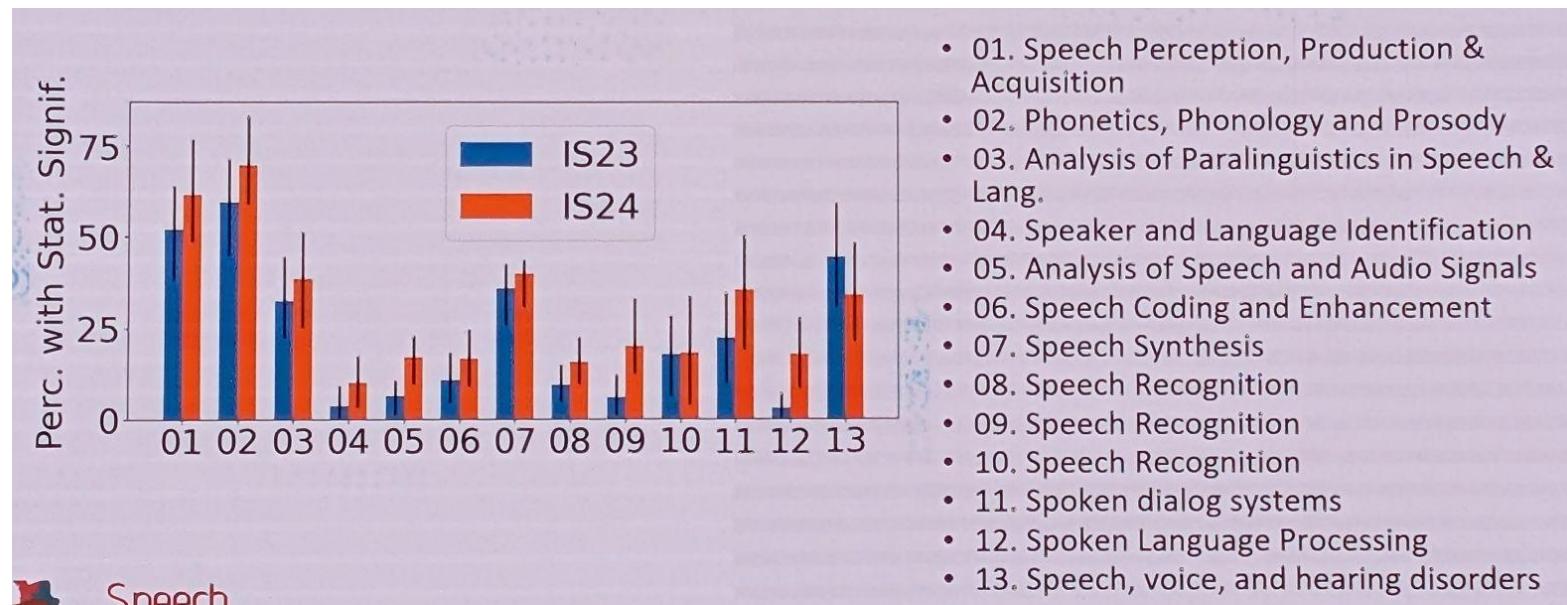
「統計的分析」の統計的分析：年ごとの傾向

- 提出時のチェック項目に「実験結果の統計的有意性または信頼区間」を追加
- 統計的分析のある論文の割合を調査



「統計的分析」の統計的分析：分野ごとの傾向

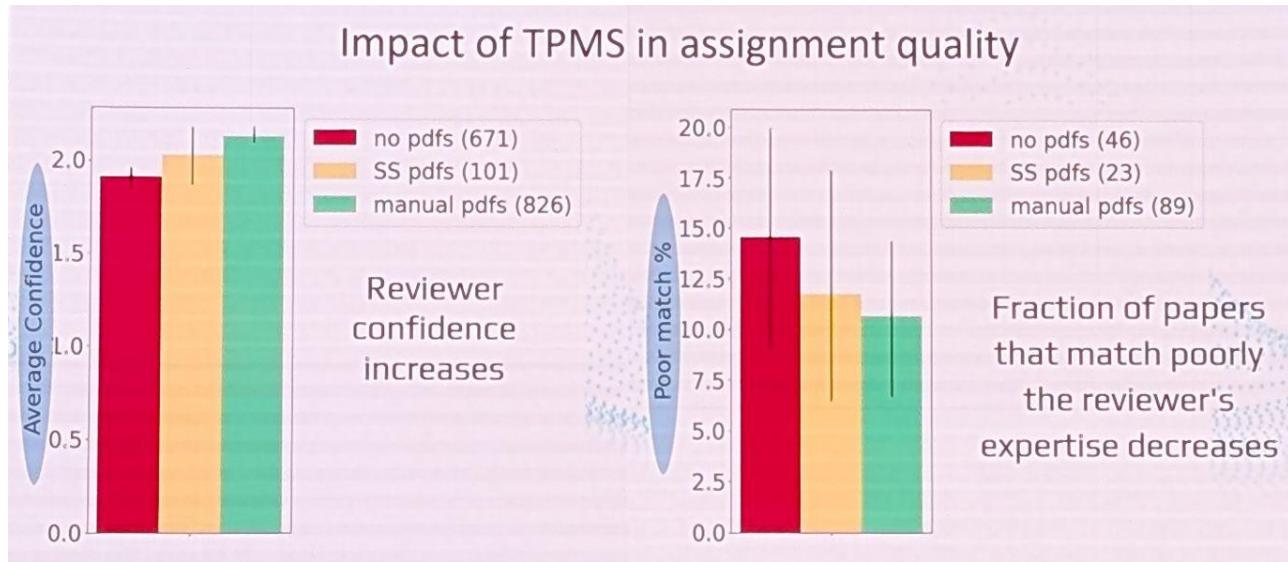
- 分野別で見ても統計的分析は増加した傾向.... ?



Opening Ceremony より引用

自動化される学会運営：査読マッチング最適化

- Toronto Paper Matching System による自動マッチング + 手動修正



Opening Ceremony より引用

自動化される学会運営：セッション構成自動化

- SPECTER2 を用いて **タイトル+概要の埋め込みベクトル**を作成
- 埋め込みに対し **クラスタリング**を実施
 - 14エリアでそれぞれクラスタリングを実行
 - 各クラスタは最大6論文まで構成される
- 昨年のINTERSPEECHとの比較によって **タイトルを提案**
- **最終決定はArea char**によって行われる
 - 5/23-24にTechnical Program Committeeが一同に集結
 - 対面でSessionを組むことで綿密な連携が可能に
- Interspeechの“STRICT RULE”: **口頭とポスターは等価**

採択論文の傾向（あくまで主観です）

- ICASSP 2024よりも「機械学習してます！」な論文は少ない傾向... ?
 - 「アーキテクチャ改良しました！」のポスターが並んでるセッションは少なかった印象
 - Anonymy Period の影響？統計的分析の影響？
 - もしくは、同セッションの論文類似度が高すぎたために掲載されていない可能性?
 - 最終日にDiffusionが偏りすぎたと日本の音声研究者で話題に
 - NLPによるセッション構成のサイドエフェクト... ?
- LLMは「一定数見るかな」程度
 - 活用している論文はもちろんあったが、LLM一色という印象も受けず

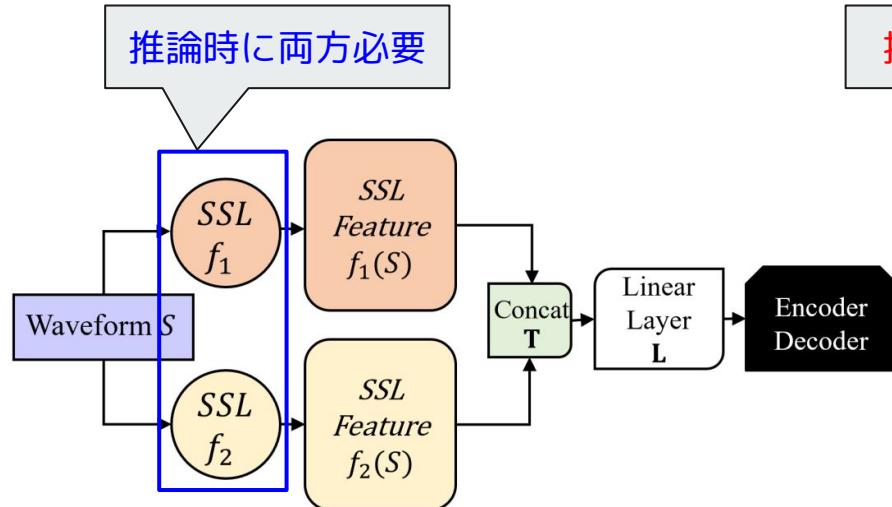
論文紹介：EFFUSE（概要）

- タイトル
 - EFFUSE: Efficient Self-Supervised Feature Fusion for E2E ASR in Low Resource and Multilingual Scenarios
- 著者
 - Tejes Srivastava (シカゴ大, CMU) , Jiatong Shi, William Chen, Shinji Watanabe (CMU)
- 概要
 - 複数SSLモデルのfeature fusionを、単一SSLモデル+fusion特徴量予測で代替する
- 選んだ理由
 - Best Paper 2件のうち1件
 - 実用上の意義が大きく、SSLモデル理解の観点でも興味深い

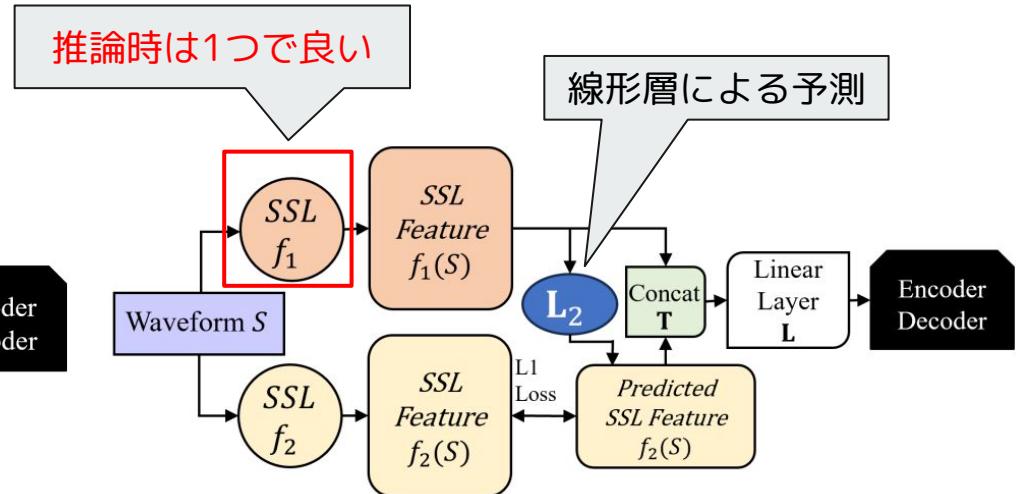
論文紹介：EFFUSE（背景）

- 音声認識（ASR）タスクで自己教師あり学習（SSL）モデルが高い性能を達成
 - HuBERT, WavLM, Wav2Vec2.0,
 - 特に low-resource や multilingual の文脈で効果的
- ドメインギャップによって十分な性能を発揮できない
 - 主要なSSLモデルは英語コーパスで学習されており、十分に汎化していない
 - 他の言語や異なる録音環境へ転用すると性能が低下する
- 従来手法：複数SSLモデルのfeature fusion
- 課題：SSLの推論コストがかさんでしまう
 - RTFの低下に繋がる

論文紹介：EFFUSE（提案手法）



Step1. feature fusionを学習（従来法）



Step2. SSL特徴量予測モデルを学習

論文紹介：EFFUSE（予備実験）

- TotonacコーパスでASRモデルを学習
 - ASRモデルはSSL各層出力の重み付き和を利用
- ある特徴量からの線形予測によって別の特徴量を予測し、相関を調査

表1：特徴量予測の相関

		Target			
		MFCC	HuBERT	WavLM	Wav2Vec 2.0
Input	MFCC	×	0.06	0.07	0.08
	HuBERT	0.67	×	0.71	0.60
	WavLM	0.66	0.70	×	0.59
	Wav2Vec2.0	0.68	0.63	0.65	×

MFCCからの
予測は困難

SSLから別のSSLは
予測可能

論文紹介：EFFUSE（実験結果1）

表2：音声認識エラー率

- low-resource languageでの検証
- 実験条件
 - 上半分：単一言語SSL
 - 下半分：多言語SSL
- いずれでも提案手法はTopline (future fussion) に匹敵する性能を達成
- RTFの観点ではToplineを凌駕している

Model	YM		Totonac	
	WER	CER	WER	CER
HuBERT (H)	21.8	10.5	55.0	23.2
WavLM (WL)	21.0	10.4	52.3	21.7
Wav2Vec 2.0 (WV)	20.5	9.9	56.4	26.3
H+WL+WV [Topline]	20.3	9.6	45.2	14.5
H→WV+WL [Proposed]	20.6	9.9	49.1	15.9
WL→H+WV [Proposed]	20.2	9.6	46.9	14.7
WV→H+WL [Proposed]	20.3	9.8	46.2	15.2
MMS (M)	22.6	10.4	48.2	16.6
XLS-R (X)	23.0	10.4	44.9	14.7
M+X [Topline]	24.1	10.7	43.8	14.0
M→X [Proposed]	21.8	9.8	46.2	15.4
X→M [Proposed]	22.6	10.1	44.6	13.9

論文紹介：EFFUSE（実験結果2）

- 多言語設定での実験を実施
- 要点をまとめると
 - リソース制約下（10分データセット）における実用性
 - タスク汎化性：ASR, LID, etc…

表3：多言語設定の評価

SSL	Info	Monolingual ASR		Multilingual ASR		LID Normal ACC ↑	Multilingual ASR + LID			SUPERB _s ↑
		CER ↓		Normal CER ↓	Few-shot CER ↓		Normal ACC ↑	CER ↓	Few-shot CER ↓	
MMS (M)	Baseline	33.8 / 30.5		29.9 / 24.6	35.0 / 35.4	62.3 / 84.3	73.8 / 88.9	29.5 / 24.5	34.9 / 35.4	1051.7 / 1029.2
XLS-R (X)	Baseline	39.5 / 30.5		28.9 / 21.6	41.4 / 39.1	65.4 / 87.2	76.9 / 90.4	28.7 / 22.0	41.3 / 38.3	957.6 / 1018.1
MMS+XLS-R (M+X)	Fusion [Topline]	35.6 / 27.9		27.0 / 20.6	38.9 / 36.6	81.3 / 86.3	78.1 / 91.9	26.8 / 20.5	38.7 / 36.9	1095.2 / 1068.4
MMS→XLS-R (M→X)	Prediction [Proposed]	38.9 / 29.4		26.5 / 22.3	33.8 / 35.4	83.4 / 84.9	82.9 / 92.1	26.3 / 21.6	33.8 / 34.0	1152.3 / 1066.5
XLS-R→MMS (X→M)	Prediction [Proposed]	38.6 / 28.9		28.3 / 21.3	41.0 / 39.7	83.9 / 90.4	76.8 / 91.0	28.1 / 21.0	39.3 / 39.1	1054.8 / 1037.0

論文紹介：EFFUSE（まとめ）

- 複数SSLモデルのfeature fusionを、単一SSLモデル+fusion特徴量予測で代替
- 実用上のメリット
 - 従来のfeature fusion手法に対しRTFやパラメタ効率を改善
 - 低リソース言語でのASR性能向上に寄与
 - 多言語設定においても有効
- (個人的に)興味深い点
 - SSLモデルの特徴量間には共通の構造性があることが示唆？

次回の開催予定：オランダ（ロッテルダム）

Flying to the Netherlands

ここ



➤ Flying to the Netherlands

Amsterdam Airport Schiphol (AMS): Also known simply as **Schiphol Airport**, it is the main airport of the Netherlands. Travelling to Rotterdam Central Station by train takes between 35-50 minutes.

Rotterdam The Hague Airport (RTM): It is a minor international airport that serves both the cities of Rotterdam and The Hague. It is located to the North of Rotterdam. Travelling to Rotterdam Central Station takes around 25 minutes.

Eindhoven Airport (EIN): It is an international airport located 7.6 KM West of Eindhoven. Flights to here tend to be cheaper. The train to Rotterdam Central Station takes around 1h:30m-2 hours.



次回の開催予定：日程表

Deadline	Date
Challenges deadline	October 14, 2024
Special sessions deadline	November 18, 2024
Notification of pre-selection SS and challenges	November 8, 2024
Tutorial submission deadline	January 25, 2025
Paper submission deadline	February 12, 2025
Show & Tell submission deadline	April 2, 2025
TPC meeting	May 7-8, 2025
Paper acceptance notification	May 21, 2025
Interspeech conference	August 17-21, 2025

締め切りまで
あと70+日