arXiv:2603.06327v1 [eess.AS] 6 Mar 2026

自闭症与非自闭症儿童语音的分类：芬兰语、法语和斯洛伐克语的跨语言研究

Sofoklis Kakouros[1]*,Ida-Lotta Myllyl¨a*[2]

[1]坦佩雷大学信号处理研究中心言语与认知研究组[2]赫尔辛基大学语言系

sofoklis.kakouros@tuni.fi,ida-lotta.myllyla@helsinki.fi

摘要

我们开展了一项针对说芬兰语、法语和斯洛伐克语的自闭症与非自闭症儿童语音的跨语言研究。我们将监督分类与语言内及跨语料库迁移实验相结合，以评估语言内和跨语言的分类性能，并探究哪些声学线索是语言特异的、哪些是语言通用的。我们使用大量声学韵律特征，将说话人级分类基准作为分析工具，而非追求最先进的性能。

语言内模型经说话人级交叉验证评估后，结果并不一致。芬兰语模型表现最佳（准确率0.84，F1分数0.88），其次是斯洛伐克语（准确率0.63，F1分数0.68）和法语（准确率0.68，F1分数0.56）。随后，我们检验了跨语言泛化。在全部合并语料上训练的模型总体准确率为0.61，F1分数为0.68。留一语料库实验用于检验向未见语言的迁移，在斯洛伐克语（F1分数0.70）和芬兰语（F1分数0.78）上取得中等成功，但向法语的迁移较差（F1分数0.42）。跨语言的特征重要性分析表明，自闭症的声学标记部分共享，但并非完全语言不变。

这些发现表明，某些与自闭症相关的语音线索能够推广到类型学差异显著的语言中，但要构建稳健的跨语言分类器，可能需要采用语言感知型建模，并确保录音条件更加一致。

索引词：韵律、儿童语音、自闭症、芬兰语、分类

1. 引言

自闭症是一种神经发育状况，其特征是社会沟通和互动方面存在差异 [1]。务实语言和言语可能偏离预期的会话规范 [2], , 且已有研究指出，自闭症说话人的韵律与非自闭症同龄人有所不同 [3]。自闭症人群中说话人之间的差异以及韵律特征的异质性十分显著，但证据表明，音高特征是常见的偏离特征 [4, 3]。音高可能变化特别大或过于夸张，但也可能呈现平坦的音高；这表明，韵律极端表现是自闭症韵律的显著特征 [5, 3, 6]。其他韵律特征差异包括强度控制、非典型停顿、时长韵律以及嗓音质量方面的异常 [7, 4, 3]。越来越多的实证证据表明，自闭症中的韵律存在差异；研究者已在多种语言中以及跨语言层面识别出与自闭症相关的韵律特征 [3, 8, 9]。通过自动分类方法，研究者利用与基频、强度以及时间和频谱特征相关的多种声学特征，从语音信号中检测出自闭症 [8, 10, 11]。

在本研究中，我们采用基于韵律特征训练的监督模型 [12] ，对单一语言内及跨语言的自闭症言语和非自闭症言语进行分类。通过比较模型在单语和跨语言设置下的表现，并分析两种条件下的特征重要性，我们识别出可能依赖于语言的与自闭症相关的韵律线索，以及可能跨语言共有的线索。

2. 材料

我们的实验使用了三个数据集，分别对应所研究的三种语言：芬兰语、法语和斯洛伐克语。下文各小节将介绍数据及预处理方法（数据概况另见表1）。

2.1. 芬兰语数据

芬兰语语音数据采集自六名以芬兰语为母语的自闭症男性，年龄为11至13岁，他们被诊断为阿斯伯格综合征（ICD-10 [13]）；该诊断现已归入自闭症谱系障碍（ASD）诊断之下。对照组（典型发育，TD）数据采集自六名年龄匹配、非自闭症、以芬兰语为母语且未报告发育问题的男性。

自闭症参与者在芬兰赫尔辛基大都市区一家医院开展的社交技能与沟通干预活动中，参与小组自发讨论时接受了录音。三名参与者参加了一场活动，另外三名参加了另一场。每场活动持续约两小时，包括自由讨论和游戏。活动由两名医疗专业人员主持。

对照组在赫尔辛基大都市区一所学校参加了由教师主持的类似讨论，录音在此过程中进行。这种对话情境是为研究目的而设置的，内容包括自由讨论。

两组参与者均根据各自地点的招募情况选取。所有活动均在安静的环境中录制，每位说话人都使用便携式录音机和头戴式麦克风，录制的音频为16位单声道，采样率为44.1千赫。

2.2. 法语数据

法语语音数据采集自六名以法语为母语的自闭症男性和三名非自闭症对照者，年龄均为11至13岁。自闭症参与者在言语治疗诊所参加自发的小组活动时进行了录音；这些活动旨在提升社交技能和沟通能力。两名参与者参加了一场活动，另四名参加了另一场。每场活动持续约一小时，同样包含自由讨论和游戏。活动由两名言语治疗师主持。三名说话人参加的那场活动在瑞士日内瓦录制，另一场则在瑞士奥尔布录制。

对照组的录音是在瑞士洛桑进行的成人主持讨论中采集的。这种对话情境专为研究而设计，同样包含自由讨论。两组参与者均根据各自所在地的人员可用情况选取。所有活动均在安静环境中录制，每位说话人使用便携式录音机和头戴式麦克风，采集16位单声道音频，采样率为44.1千赫。

2.3. 斯洛伐克语数据

本研究使用的斯洛伐克语语音数据来自斯洛伐克自闭症与非自闭症儿童语音语料库（SANACS）[14]。该语料库由儿童（自闭症或非自闭症）与一名非自闭症成人实验员之间合作完成的任务导向型对话录音组成，所有参与者均为母语说话人。数据涵盖67名6至12岁斯洛伐克儿童的语音（平均年龄为9.2岁）。其中，37名说话人患有自闭症（7名女孩、30名男孩），30名为非自闭症儿童（7名女孩、23名男孩）。每名儿童都与同一位实验员互动。所有自闭症儿童均被诊断为自闭症谱系障碍（ASD）。研究使用地图任务来采集该语料库。大多数任务包含六轮：一轮练习，以及两轮由实验员描述、儿童跟随；随后再进行一轮练习和两轮角色互换的任务。所有场次均使用便携式录音机和便携式麦克风录制立体声音频，其中声道1为实验员，声道2为儿童。该语料库由斯洛伐克科学院信息学研究所与夸美纽斯大学自闭症学术研究中心合作录制，两家机构均位于斯洛伐克布拉迪斯拉发。

2.4. 数据预处理

所有成年场次主持人的语音数据均未纳入分析。对于斯洛伐克语数据，儿童语音已从立体声音频中提取并转换为单声道。所有儿童说话人的语音数据均被切分为IPU（语段间停顿单元），即由同一说话人产生、停顿至少为200毫秒的语音片段。切分过程通过强度阈值自动完成，并经人工验证。在分析之前，含有重叠语音的IPU以及仅由非词汇性发声构成的IPU均被剔除。在此阶段，斯洛伐克语数据集中有一名说话人被排除，因为其数据中包含大量重叠语音，且IPU仅由非词汇性犹豫标记构成。

3. 方法

3.1. 特征提取

所有语音录音均使用 openSMILE 工具包进行处理。该开源工具包用于音频特征提取，并采用 eGeMAPS 话语级统计泛函配置[15],，为每个语段间停顿单元（IPU）生成一个88维特征向量。此配置将选定的低级描述符与其统计泛函相结合，涵盖基频（$f0$）、强度、频谱特征和嗓音质量指标（例如基频微扰、振幅微扰）。提取的特征全面表征韵律动态和嗓音质量特征，捕捉音高与能量分布、频谱平衡及时间变异性。

特征输出被汇总到一个数据集中，其中还包含每个语调单位的元数据（例如说话人 ID、组别、语言、性别和年龄）。这些元数据字段与 88 维声学韵律特征向量分开存储。

语言	自闭症谱系障碍（话语/编号）	典型发育（话语/编号）	平均时长（秒）	总计（话语/编号）
芬兰语	1323 / 6	249 / 6	2.11	1572 / 12
法语	884 / 6	659 / 3	1.79	1543 / 9
斯洛伐克语	2867 / 37	2230 / 29	2.36	5097 / 66
ALL	5074 / 49	3138 / 38	2.20	8212 / 87

表1：各语言的话语（*utt.*）数量、每个诊断组中的说话人标识符数量，以及平均录音时长。tor，并根据文件名与每个语调单位相关联。

3.2 分类基准

为更好地理解跨语言差异以及自闭症言语与非自闭症言语之间的区别，我们采用了一个简单的监督分类框架，以评估语言内和跨语言泛化。具体而言，我们使用通过 openSMILE 提取的 88 维声学特征集，训练了简单的分类器，并采用 XGBoost[16] 和随机森林模型来区分自闭症与非自闭症说话人的样本。语言内实验评估了在同一种语言上训练和测试的模型捕捉语言特异声学标记的能力；跨语言实验则考察在一种语言（或部分语言）上训练的模型泛化到另一种语言的程度，从而探究所识别线索在不同语言情境中的稳健性和可迁移性。

需要强调的是，我们的目标并不是在这些语言的自闭症谱系障碍（ASD）与典型发育（TD）分类任务中确立最先进的性能。相反，我们有意采用简单且可解释的分类器，并将其应用于一组成熟的声学韵律特征。这一选择使我们能够优先考虑透明度和解释力，而非原始预测准确率，并将这些模型作为分析工具，以更清晰地理解特定韵律和声学特征如何与不同语言中的诊断组差异相关。

3.3 特征重要性

在特征重要性分析中，我们采用了多样化的基于模型和模型无关的方法，以获得稳健且可解释的估计结果。具体而言，我们采用了基于树的集成方法，包括 XGBoost、随机森林和单棵决策树，根据分裂增益和不纯度降低计算重要性评分。为补充这些特定于模型的指标，并提供理论依据更充分的归因分析，我们应用 TreeSHAP [17] 量化每个特征对模型输出的边际贡献；此外，我们还使用基于置换的特征重要性对这些发现进行了进一步验证。该方法通过评估随机打乱单个特征对预测性能的影响来衡量其重要性。所有分析均分别针对每种语言进行，并在合并后的多语言数据集上进行，从而使我们能够考察语言特定模式以及在全球范围内稳定的预测变量。

4. 实验

4.1 分类

4.1.1. 语言内分析

对于语言内设置，我们针对每种语言（斯洛伐克语、芬兰语和法语）分别训练模型。为了真实估计模型对未见过的个体的泛化能力，我们在说话人层面进行交叉验证，确保——

自闭症谱系障碍和典型发育说话人同时出现在训练集和测试集中。在实际操作中，我们基于不重复的受试者标识符定义折，并确保同一说话人的所有话语在给定折中要么被分配到训练集，要么被分配到测试集，但绝不会同时分配到两者。我们采用名义上的五折设置；当说话人数或类别分布形成限制时，会自动调整有效折数，以维持有效划分，同时至少保留两折。关键在于，模型评估始终反映模型对训练期间未见过的说话人的表现。

由于我们的数据集在性别方面严重失衡（大多数录音来自男性参与者），因此无法可靠地实施性别平衡划分。我们仅在唯一受试者标识符层面进行随机划分。分类器采用类别加权训练，以部分补偿自闭症谱系障碍（ASD）与典型发育（TD）标签不平衡（另见表1）。然而，所有数据集的这种不平衡都无法得到完全纠正，因为芬兰语数据的偏斜尤为严重。随后，我们针对每种语言分别运行随机森林和XGBoost，并报告按折平均的标准准确率和F1分数（均值和标准差）。

4.1.2. 跨语言分析

为考察跨语言泛化，我们采用两种互补方案：（1）多语言汇总训练；（2）留一语料库评估（LOCO）。在第一种方案中，我们将三种语言的数据汇总为一个数据集，并再次在说话人层面进行5折交叉验证（同样确保说话人不会同时出现在训练折和测试折中）。该方案旨在考察单一模型能否利用规模更大、更多样化的训练集，以及接触多种语言是否能使分类器在面对语言特异的声学特异性时保持稳健。它也提供了与语言特异基线的比较依据：如果汇总模型表现相当或更好，则表明跨语言信息有益而非有害。

在第二种跨语言方案中，我们进行留一语料库（LOCO）评估，以更直接地检验跨语言迁移。在这里，我们将每种语言视为一个独立语料库。每次迭代时，我们完全留出一种语言用于测试，并使用其余两种语言训练模型。例如，我们可以使用斯洛伐克语和芬兰语进行训练、以法语进行测试；然后使用斯洛伐克语和法语进行训练、以芬兰语进行测试，依此类推，直到每种语言都作为未见过的测试集使用一次。与合并数据的设置不同，LOCO 中训练集和测试集之间不会混合语言：模型是在训练过程中从未接触过的语言上进行评估的。这种设置探讨的是一个不同的问题：区分自闭症谱系障碍（ASD）与典型发育（TD）的声学韵律标记，在不同语言之间有多大程度的稳定性，以至于在某些语言上训练的模型能够泛化到一种全新的语言。若模型在 LOCO 设置下表现良好，则表明 eGeMAPS特征集所捕捉到的至少部分 ASD 与 TD 特征具有语言无关性；这也意味着，在谨慎使用的前提下，基于资源较丰富的语言训练的模型或可应用于数据很少甚至没有数据的语言。

4.2. 特征重要性

4.2.1. 语言内分析

对于特征重要性，我们采用基于共识的方法，以更好地捕捉最相关的特征。特征重要性通过决策树、随机森林、XG-Boost、TreeSHAP 和置换重要性进行估算（另见 [18]）。对于每个特征，我们汇总这些方法得出的排名，计算其平均排名（及最佳排名）。随后，我们选出至少被两种方法判定为重要的特征，按其平均排名/最佳排名排序，并为每个范围和语言确定一组共识性的前五项特征。

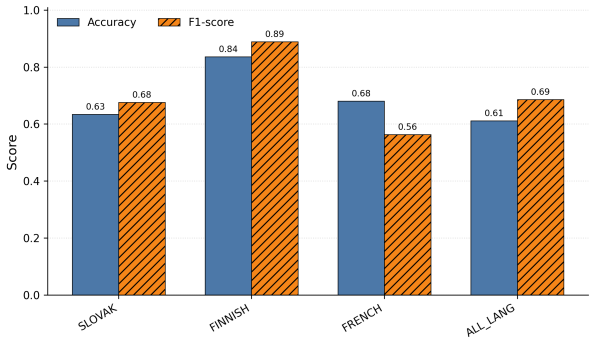


图 1：*XGBoost* 模型按语言及总体划分的准确率和 *F1*分数。

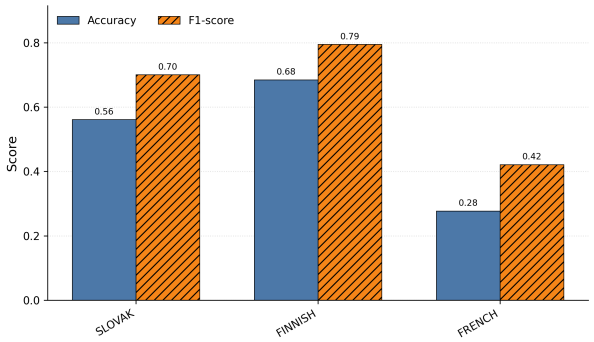


图 2：*XGBoost* 模型按语言划分的 *LOCO* 准确率和 *F1*分数。

4.2.2. 跨语言分析

在跨语言分析中，我们采用与语言内分析相同的程序，但将其应用范围扩展至包含所有三种待研究语言的完整数据集。这样，我们便能评估在单种语言中观察到的模式，在跨语言联合考察数据时是否同样成立。

5. 结果与讨论

在接下来的小节中，我们将针对每项语言内和跨语言分析，呈现单个模型（XGBoost，因为它与随机森林的差异通常较小）的分类结果，以及相应的特征重要性。

5.1. 语言内分析

语言内分类结果表明，简单的分类设置能够区分自闭症谱系障碍组和典型发育组，不过不同语言的表现有所差异，如图1所示。总体表现最佳的是芬兰语（准确率84%，F1分数89%），其次是斯洛伐克语（准确率63%，F1分数68%）和法语（准确率68%，F1分数56%）。不过，解读芬兰语结果时应当谨慎。首先，数据集的类别分布严重不平衡，自闭症谱系障碍组的

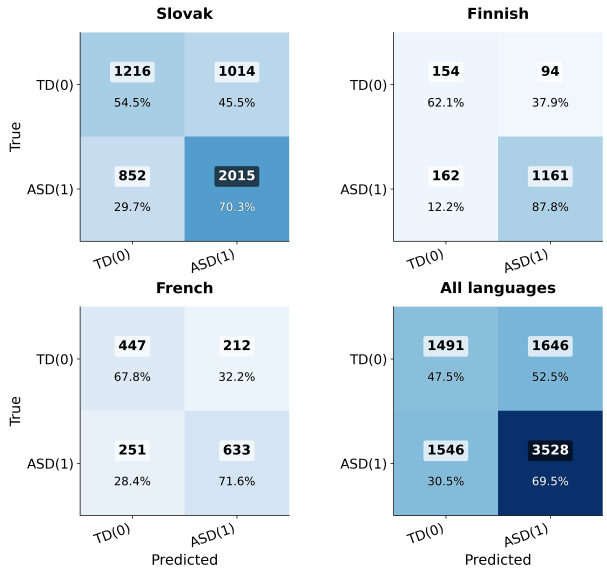


图 3： *XGBoost* 模型按语言及总体的混淆矩阵。

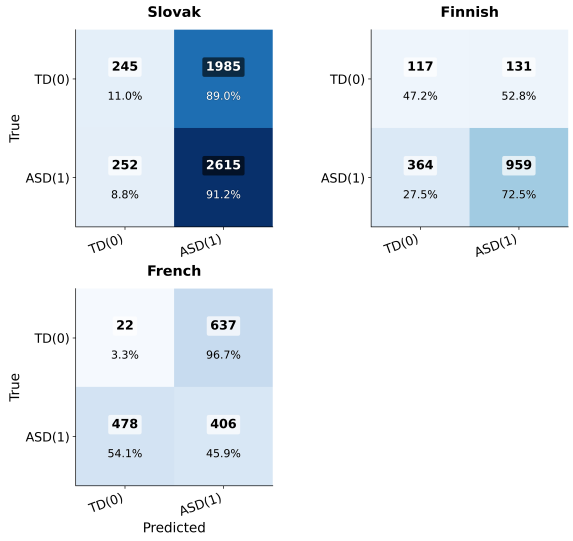


图 4： *XGBoost* 模型按语言进行留一语言交叉验证的混淆矩阵。

样本数量远多于另一组（见表1），这也反映在混淆矩阵中（图3）。其次，我们观察到芬兰语中自闭症谱系障碍组和典型发育组的说话风格存在定性差异，这可能是因为自闭症谱系障碍组与临床医生更为熟悉；这种熟悉程度似乎会促使他们使用更生动、更具表现力的语言，而非简单回应或简短陈述。总体而言，法语和斯洛伐克语的结果似乎更可靠、更稳健。斯洛伐克语的结果尤其如此，其类别分布大致均衡，且可用的数据集规模较大、样本多样性较高。

特征重要性结果见图 5。在芬兰语、斯洛伐克语和法语中，基频分布（音高水平和范围）始终是区分自闭症谱系障碍与典型发育语音的重要特征，而次要线索因语言而异：芬兰语主要依赖频谱倾斜度（例如，参见 [19]）和嗓音质量，斯洛伐克语依赖整体频谱形状和动态特征，法语则依赖较高共振峰结构和整体强度。

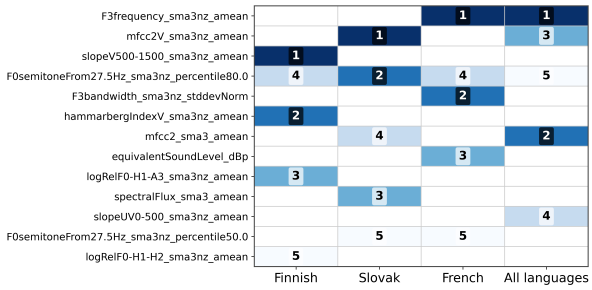


图 5： 各语言及所有语言的共识特征重要性。

5.2. 跨语言分析

如图 1、2、3 和 4 所示，我们的跨语言分析表明，在当前简单的设置下，模型的泛化能力有限。汇集所有语言的数据后，模型在各数据划分上的平均表现为准确率 61%、F1分数 69%。然而，在留一语言交叉验证设置下，模型表现因语言而异：芬兰语的结果最佳（准确率 68%、F1分数 79%），其次是斯洛伐克语（准确率 56%、F1分数 70%）和法语（准确率 28%、F1分数 42%）。总体而言，这些结果表明，各语言确实共享一些与任务相关的信息，但与特征重要性分析的结果一致，它们可能依赖不同的特征。还需注意，留一语言交叉验证设置下准确率与 F1分数之间的显著差异是由类别不平衡导致的（见图 4）。

就特征重要性而言，对于所有语言的合并数据，图 5 中的特征表明，音段特征（元音和辅音）与音高属性都以相对语言通用的方式发挥作用。总体而言，针对各语言的分析显示出略有不同的侧重点，但合并模型凸显了一组共同的、语言通用的韵律及频谱与音段线索（与音高和频谱形状相关的特征），这些线索能够持续地区分自闭症谱系障碍语音与典型语音。

6. 结论

总体而言，我们的研究结果表明，基于韵律特征训练的监督模型能够较好地区分同一语言中的自闭症谱系障碍（ASD）语音与典型发育（TD）语音，在跨语言区分方面的效果则较为有限。这支持了与自闭症相关的韵律线索既具有语言特异性，也具有语言通用性这一观点。跨语言来看，汇总分析和留一语言交叉验证（LOCO）分析显示，模型的泛化能力有限，且不同语言间存在显著差异。这表明，尽管某些与任务相关的信息为不同语言所共享，但这些信息并未编码于完全重叠的声学韵律模式中。与此一致，特征重要性分析发现，F0分布是一种稳健的语言通用线索，同时还发现了语言特异的频谱、声音质量和音段特征。这表明，自闭症谱系障碍语音的韵律标记部分具有通用性，但也受到语言特异性特征的影响。

7. 致谢

SK 得到了 Kone 基金会资助的 L-SCALE 项目的支持。ILM 获得了芬兰研究理事会的资助，资助编号为 362906。作者谨此感谢芬兰 CSC – IT 科学中心提供计算资源。

8. 参考文献

[1] C. Lord、M. Elsabbagh、G. Baird 和 J. Veenstra-Vanderweele，《自闭症谱系障碍》，《柳叶刀》，第 392 卷，第 10146 期，第 508–520 页，2018 年。[2] M. Grice、S. Wehrle、M. Kr¨uger、M. Spaniol、F. Cangemi 和 K. Vogeley，《自闭症谱系障碍中的语言韵律——概述》，《语言与语言学指南》，第 17 卷，第 5 期，文章编号 e12498，2023 年。[3] R. Fusaroli、R. Grossman、N. Bilenberg、C. Cantio、J. R. M. Jepsen 和 E. Weed，《迈向自闭症语音标记的累积性科学：基于跨语言元分析的美国和丹麦自闭症儿童声学标记研究》，《自闭症研究》，第 15 卷，第 4 期，第 653–664 页，2022 年。[4] R. Fusaroli、A. Lambrechts、D. Bang、D. M. Bowler 和 S. B. Gaigg，《声音是自闭症谱系障碍的标记吗？系统综述与元分析》，《自闭症研究》，第 10 卷，第 3 期，第 384–407 页，2017 年。[5] M. Wiklund 和 M. Vainio，《团体治疗课程数据中自闭症芬兰语和法语男孩语音的音高相关特征》，《比较语用学研究》，第 45 页，2019 年。[6] D. Okuizumi、K. Terada、A. Ishii、Y. Ohmoto、H. Shimizu、A. Imamura、R. Iwanaga 和 H. Kumazaki，《5 岁自闭症谱系障碍儿童情绪表达的声学特征》，《精神病学前沿》，第 16 卷，文章编号 1444675，2025 年。[7] J. Trayvick、S. B. Barkley、A. McGowan、A. Srivastava、A. W. Peters、G. A. Cecchi、J. H. Foss-Feig 和 C. M. Corcoran，《自闭症中的言语和语言模式：将自然语言处理作为研究与临床工具》，《精神病学研究》，第 340 卷，文章编号 116109，2024 年。[8] J. C. Lau、S. Patel、X. Kang、K. Nayar、G. E. Martin、J. Choy、P. C. Wong 和 M. Losh，《自闭症言语韵律差异的跨语言模式：一项机器学习研究》，《公共科学图书馆·综合》，第 17 卷，第 6 期，文章编号 e0269637，2022 年。[9] S. Z. Asghari、S. Farashi、S. Bashirian 和 E. Jenabi，《自闭症谱系障碍人士独特的韵律特征：系统综述与元分析》，《科学报告》，第 11 卷，第 1 期，文章编号 23093，2021 年。[10] N. A. Chi、P. Washington、A. Kline、A. Husic、C. Hou、C. He、K. Dunlap 和 D. P. Wall，《根据众包半结构化语音录音对自闭症进行分类：机器学习模型比较研究》，《*JMIR* 儿科与育儿》，第 5 卷，第 2 期，文章编号 e35406，2022 年。[11] A. Mohanta 和 V. K. Mittal，《使用声学特征分析并分类自闭症谱系障碍儿童的语音》，《计算机语音与语言》，第 72 卷，文章编号 101287，2022 年。[12] S. Kakouros 和 K. Hiovain-Asikainen，《利用自监督语音模型识别北萨米方言》，载于*Interspeech 2023*，2023 年，第 5306–5310 页。[13] 世界卫生组织，*ICD-10*：《疾病和有关健康问题的国际统计分类》，第 10 版。日内瓦：世界卫生组织，2004 年。[14] J. Kruyt、R. Sabo、K. Pol´onyiov´a、D. Ostatn´ıkov´a 和 ˇ S. Beˇnuˇs，《斯洛伐克自闭症与非自闭症儿童语音语料库：以任务为导向的儿童与成人互动》，载于*2024* 年计算语言学、语言资源与评测联合国际会议论文集（*LREC-COLING 2024*），2024 年，第 16 094–16 099 页。[15] F. Eyben、K. R. Scherer、B. W. Schuller、J. Sundberg、E. Andr´e、C. Busso、L. Y. Devillers、J. Epps、P. Laukka、S. S. Narayanan 等，《用于语音研究和情感计算的日内瓦精简声学参数集（GeMAPS）》，《*IEEE* 情感计算汇刊》，第 7 卷，第 2 期，第 190–202 页，2015 年。[16] T. Chen，《XGBoost：一种可扩展的树提升系统》，康奈尔大学，2016 年。

[17] S. M. Lundberg、G. Erion、H. Chen、A. DeGrave、J. M. Prutkin、B. Nair、R. Katz、J. Himmelfarb、N. Bansal与S.-I. Lee，“利用面向树的可解释人工智能从局部解释走向全局理解”，《自然·机器智能》，第2卷，第1期，第56–67页，2020年。[18] S. Kakouros与H. Chen，“听起来像赢家？赛后采访中的韵律差异”，载于*Interspeech 2025*，2025年，第4188–4192页。[19] S. Kakouros、O. R¨as¨anen与P. Alku，“语音中句子显著性的谱倾斜度量比较——维度与不利噪声条件的影响”，《语音通信》，第103卷，第11–26页，2018年。