arXiv:2602.04247v1 [cs.CL] 4 Feb 2026

DementiaBank-Emotion：阿尔茨海默病语音多标注者情绪标注语料库（版本 1.0）

Cheonkam Jeong[1], [7], [*]、Jessica Liao[2], 、Audrey Lu[2],、Yutong Song[2],、Christopher Rashidian[2],、Donna Krogh[3], 、Erik Krogh[3], 、Mahkameh Rasouli[1],、Jung-Ah Lee[1],、Nikil Dutt[2],、Lisa Gibbs[6], 、David Sultzer[4], 、Julie Rousseau[6],、Jocelyn Ludlow[1],、Margaret Galvez[2],、Alexander Nuth[8], 、Chet Khay[5],、Sabine Brunswicker[7], [†]，Adeline Nyamathi[1], [†]

[1]加利福尼亚大学欧文分校苏·格罗斯与比尔·格罗斯护理学院[2]加利福尼亚大学欧文分校唐纳德·布伦信息与计算机科学学院，美国[3]Smart Forward，兰乔帕洛斯弗迪斯，美国[4]加利福尼亚大学欧文分校精神病学与人类行为系，美国[5]Amore老年生活中心，拉古纳尼古尔，美国[6]加利福尼亚大学欧文分校医学院[7]普渡大学，西拉斐特，美国[8]乔治·华盛顿大学，华盛顿哥伦比亚特区，美国

摘要

我们介绍了 DementiaBank-Emotion，这是首个针对阿尔茨海默病（AD）语音的多标注者情绪标注语料库。我们对 108 位说话人的 1,492 个话语进行了标注，涵盖 Ekman 的六种基本情绪和中性，发现阿尔茨海默病患者表达的非中性情绪显著更多（16.9%），高于健康对照组（5.7%； $p < .001$）。探索性声学分析显示，两组可能存在分离：对照组说话人表达悲伤时，基频调制幅度较大（相较基线为 $\Delta = -3.45$ 个半音），而阿尔茨海默病说话人的变化极小（$\Delta = +0.11$ 个半音；交互作用 $p = .023$）。不过，这一发现基于有限样本（悲伤： n=5 名对照组说话人， n=15 名阿尔茨海默病患者），仍需重复验证。在阿尔茨海默病语音中，响度可区分不同情绪类别，表明情绪—韵律映射得到部分保留。我们公开发布该语料库、标注指南及校准研讨会材料，以支持临床人群情绪识别研究。

1 引言

痴呆症的特征不仅包括认知衰退，还包括语言和沟通能力的逐步受损（Mueller 等人，2018）。针对阿尔茨海默病语音分析的计算方法已取得显著进展，相关研究发现了多种语言学标记，例如词汇多样性降低、句法简化以及停顿增多（Fraser 等人，2016；Ahmed 等人，2013）。ADReSS 挑战赛（Luz 等人，2020）基于痴呆症语料库匹兹堡语料库（Becker 等人，1994年）中的声学和语言特征，建立了阿尔茨海默病自动检测的基准。

然而，阿尔茨海默病语音的情感维度仍未得到充分研究。语音情绪识别受到自然语言处理和语音处理领域的广泛关注，IEMOCAP 等数据集（Busso 等人，2008年）推动了对对话中多模态情绪识别的研究。然而，这些资源关注的是健康人群，通常采用表演式或脚本化语音。目前尚无可用于认知障碍临床人群的同类资源。

理解阿尔茨海默病语音中的情绪表达有几个重要原因。首先，情绪可能成为疾病进展的额外标志，补充现有的语言学和声学特征(Henry 等人，2009)。其次，照护者和临床医生需要准确解读阿尔茨海默病患者的情绪线索，而这些线索可能不同于典型模式。第三，具备情绪感知能力的辅助技术需要使用能够反映目标用户实际情绪表达的训练数据。

本文介绍 **DementiaBank-Emotion(v1.0)**，这是一个针对阿尔茨海默病语音的多标注者情绪标注语料库。我们的贡献如下：

1. 我们发布首个阿尔茨海默病语音情绪标注语料库：包含来自108位说话人的1,492条话语及多评分者标签（校准后的 Fleiss' κ =0.23–0.31）。

2. 我们记录了临床言语特有的标注难点，并提供校准研讨会材料，探讨如何区分保全面子的笑声和喜悦等问题。

3. 我们发现，阿尔茨海默病患者表达的更多非中性情绪多于对照组（16.9%，而

5.7%），探索性证据表明其韵律区分能力降低（声学平坦化），值得进一步研究。

4. 我们表明，响度能够区分阿尔茨海默病语音中的情绪类别，这表明情绪—韵律映射得到部分保留。

2 相关工作

2.1 自发言语中的情绪识别

IEMOCAP（Busso 等人，2008年）、MSP-IMPROV（Busso 等人，2017年）和 RAVDESS（Livingstone 和 Russo，2018）等基准数据集为语音情绪识别（SER）提供了基础资源。然而，这些数据集主要依赖来自健康个体的表演性或半脚本语音，其中通常包含典型且夸张的情绪表达。

自发临床言语，尤其是阿尔茨海默病（AD）患者的言语，带来了这些数据集未能应对的独特挑战。在阿尔茨海默病中，声学标记与情绪状态之间的映射常常受到嗓音质量病理性变化的干扰，例如情感平淡或降低 F_0 的变异性（Kreiman 等人，1993）。此外，语言焦点、语用意图和功能[1],对于识别惊讶等情绪可能至关重要，但它们在认知衰退过程中经常受到损害，因此需要比标准类别标签标注更细致的标注框架。

2.2 阿尔茨海默病语音的声学与语言学分析

DementiaBank Pitt 语料库（Becker 等人，1994年；Lanzi 等人，2023）已被广泛用于阿尔茨海默病分类，研究重点主要是思想密度、词汇多样性和言语不流畅模式（Fraser 等人，2016年；Luz 等人，2020）。尽管嗓音质量特征和 eGeMAPS 集（Eyben 等人，2016年）已成功区分阿尔茨海默病和轻度认知障碍患者与健康对照组（Themistocleous 等人，2020），但这些研究通常将声学特征视为认知状态的诊断标记，而非情绪表达的载体。

理解阿尔茨海默病语音情感层面的关键缺口依然存在。标准情感分析和自动化语音情绪识别（SER）系统往往无法考虑临床细微之处——例如，笑声可能是应对找词困难的一种机制，而非喜悦的信号（Glenn，2003年）。Chou 等人（2025年）近期探索了结合面部和眼动追踪数据与情绪预测的多模态阿尔茨海默病分类，发现仅凭情绪特征并未显著提高阿尔茨海默病检测准确率。这表明，情绪与认知状态之间的关系可能比简单的特征拼接所能捕捉到的更为微妙。我们的研究通过引入多标注者标注层来填补这一空白，该标注层明确纳入语言焦点、语用意图和临床专业知识，从而能够细粒度地分析情绪如何表达，而不仅仅是它们是否存在。

3 语料库描述

3.1 来源数据

DementiaBank Pitt 语料库（Becker 等人，1994年；Lanzi 等人，2023）包含英语参与者完成波士顿失语症诊断检查中的 Cookie Theft 图片描述任务时的音频录音和转录文本。在这项任务中，参与者需要描述一幅场景图：一名妇女在厨房里洗餐具，水从水槽溢出；孩子们从罐子里偷饼干，同时其中一个孩子坐着的凳子正倾倒。

对于本语料库（v1.0），阿尔茨海默病患者和对照组的数据均取自 ADReSS 2020 挑战赛训练集（Luz等人，2020），该训练集提供了纵向 DementiaBank 数据的横断面子集，其中阿尔茨海默病组和对照组相互匹配（每组 54 名说话人，年龄和性别均衡）。未来版本（v2.0）将纳入完整的 DementiaBank 纵向数据。

3.2 语料库统计

表1 概述了该语料库。我们标注了来自 108 名说话人的话语：54 名阿尔茨海默病患者和 54 名健康对照组，均来自 ADReSS 挑战赛数据集；该数据集在设计上使各组之间的人口统计学特征（年龄、性别）相互匹配。在情绪分析中，我们重点关注参与者（PAR）话语；研究者（INV）话语主要为中性提示语和应答语。

在总计 1,492 条 PAR话语中，615 条阿尔茨海默病话语和 731 条对照组话语获得了有效的最终情绪

[1]在此语境中，“语用”是一个一般性的描述术语，而非特定语言学理论框架内严格界定的技术术语。

	AD	对照组
说话人	54	54
会话	54	54
PAR话语	752	740
有效标签	615	731
模糊	137	9

表 1：按组别划分的语料库统计。

标签。其余 137 条阿尔茨海默病话语和 9 条对照组话语因标注者意见不一致且无法通过裁定算法解决，被标记为 模糊；这些话语未纳入情绪分布分析，但在适用情况下仍保留用于声学分析。

4 标注

4.1 标注员

标注工作由 11 名专业背景各异的评定者完成。第一轮由临床专家（四名博士级护理研究人员和一名执业护士）进行。第二轮由三名临床专家和一名商学教授参与。第三轮以及对照数据的标注由技术团队完成，团队成员包括计算机科学和理工科研究人员，并由临床专家负责人监督。这种多学科构成确保标签既反映临床实际，也满足计算分析所需的结构一致性。

4.2 标注方案

我们采用了Ekman的基本情绪类别（Ekman，1992年），并加入中性类别，共设七个标签：中性、喜悦、悲伤、恐惧、愤怒、惊讶和厌恶。为确保概念一致性，我们的标注指南纳入了Paul Ekman Group[2]提供的正式定义。我们对标注员进行培训，使其不仅能够识别典型的语音表达，还能识别与各类别相关的深层心理主题——例如，通过失落或无助等主题识别悲伤，以及将惊讶视为对Cookie Theft图片中意外视觉刺激的反应。兼顾声学特性与普遍情绪的理论定义，为解读阿尔茨海默病语音中复杂的情感信号提供了稳健的框架。

[2]https://www.paulekman.com/universal-emotions/

情绪	F0	基频变化	响度	Rate
Joy	–	High	High	中等
悲伤	Low	Low	Low	Slow
Fear	High	High	High	Fast
愤怒	Low	Low	High	Fast
厌恶	Low	Low	Low	Slow
惊讶	显著的音高变化、倒吸气			
中性	Flat, 单调			

表2：根据情绪类别预期的韵律线索，依据以下文献改编的标注指南Sobin 和 Alpert (1999)。

4.3 标注指南

最初的标注框架由首席研究员制定（其为语言学博士，专攻音系学和语音学）。该框架将Ekman的心理学情绪类别与客观声学标记相结合，例如 F_0方差、音量阈值和发声类型（Ekman，1992年；Sobin和Alpert，1999）。表2总结了各情绪类别预期的韵律线索。

4.4 校准研讨会

为解决初始标注阶段发现的分歧，我们开展了多次校准研讨会。参与者包括标注员，以及由一位精神病学教授和一家记忆照护机构主任组成的顾问小组。这些研讨会形成了一个反复迭代的反馈循环：每完成一批标注，首席研究员都会提供反馈，顾问小组则会讨论模糊案例。

这些讨论的一项主要成果是进一步明确了**默认归为中性**原则，尤其考虑到任务本身是图片描述。我们发现，即使说话人的表达平淡，标注员也常受带有情绪色彩的词汇影响（例如“偷”饼干罐）。顾问小组达成共识，优先考虑韵律线索而非词汇内容，由此制定了正式的指南 v2.0（见附录A），并独立完成了最后几轮标注。

笑声子类型。 并非所有笑声都表示喜悦。研讨会讨论区分了：(1) 愉快的笑声 伴随真切的愉悦（标注为喜悦），(2) 无助的笑声 与“可怜”等同情表达同时出现（标注为悲伤），以及 (3) 讽刺性笑声（取决于语境）。这一差异对于临床言语至关重要，

话语	标签	Cues	Spk	基频（半音）		基频方差		响度	
				eGe	spk	eGe	z	eGe	z
“他的凳子快要倒了 他 &=笑着”	Joy	愉快的笑声	234（男性，66）	24.2	−2.8	0.26	+0.27	0.66	+0.50
“嗯，可怜的母亲正在做（ 餐具 &=笑声”	Sad	无助的笑声	337（女，59岁）	33.2	+0.4	0.15	−0.69	0.38	−0.62
“他就要从那张凳子上掉下来了 too”	Surp	“凳子”上的重音	234（男性，66岁）	25.8	−1.2	0.28	+0.78	0.41	−0.57
“嗯，孩子们正在偷一个 饼干罐”	Neut	语调平淡，陈述事实	234（男性，66）	25.8	−1.2	0.18	−1.28	0.93	+1.71

表3：校准工作坊中带有一致标签、诊断线索和声学特征的示例。Spk = 说话人编号（性别、年龄）。eGe = eGeMAPS 数值（基频以 27.5 Hz 为基准的半音数表示；基频方差 = 标准差归一化；响度以宋表示）。spk/z = 说话人归一化（基频以说话人均值为基准的半音数表示；其他数值在说话人内部进行 z 评分）。声学数值在标注后计算。

因为笑声常常起到维护面子或应对困境的作用。

韵律焦点与惊讶。 遵循焦点韵律理论（Bolinger，1958；Selkirk，1995），出现显著音高变化（F_0起伏）、倒吸气，或对意料之外的元素使用强重音（例如，“他要从那张 凳子 上掉下来了”）的话语，会被标注为惊讶，如表3所示。在这些情况下，声学突出度起到语用焦点和新信息（信息焦点）标记的作用（z = +0.78）。这表明说话人的认知状态已从常规描述转变为主动“注意到”场景中显著且出乎意料的元素。

默认归为中性。 平淡的语调和客观、描述性的内容是判断为中性的主要依据。即使词语本身可能暗示某种情绪（例如“抢劫”），只要韵律表达方式平淡，标注员就会将话语标记为中性。

4.5 评分者间信度

表4 展示了各标注批次的评分者间信度。对于患者数据，Fleiss’ κ在校准工作坊就模糊案例达成共识后，从0.094（批次1，工作坊前）提高到0.313（批次2）。

Data	Fleiss’ κ	n 评定者
患者批次 1	0.094	5
患者批次 2	0.313	5
患者批次 3	0.231	5
对照组（排除 1 名）	0.254	3

表 4：各标注批次的评分者间信度。对于对照数据，由于一名评分者的标注模式存在显著差异，因此将其排除（见正文）。

对于对照数据，标注过程未进行单独的校准工作坊，因为时间有限；我们改为收集标注员的书面反馈。四名标注员中，有两名曾参与患者标注任务，另外两名标注员（L3 和 L4）是新招募的。各标注员的非中性标注率差异较大（介于 5% 至 26%）；若排除两名新标注员，评定者人数将不足，因此我们根据两两一致性分析仅排除了 L3，使剩余 3 名评定者的 κ = 0.254。

书面反馈揭示了标注策略方面的一个关键差异。经验丰富的标注员明确表示，他们会关注说话人内部的变化：L1 指出：“我根据他们典型模式的变化进行标注”，L2 则表示：“我关注不同话语之间语音音高／语速／音调的差异。”相比之下，一名新招募的标注员表示自己高度依赖明确指南（“90% 依据指南，只有 10% 依靠感知”），这表明其遵循规则与运用感知判断之间存在失衡。

这种差异表明，有效的情绪标注需要将明确标准与隐性知觉技能结合起来，尤其是识别说话人偏离基线情感的能力。校准研讨会似乎有助于促成这种结合，其作用不仅在于澄清指南，还在于通过讨论模糊案例培养共同直觉。如果缺少这样的校准，标注员可能会倾向于走向任一极端：过度依赖指南（遗漏说话人内部的细微变化），或过度依赖感知（采用个人化标准）。未来的标注工作应为所有数据安排校准环节，以建立这种平衡。

尽管这些信度值属于中等——按

常规标准——它们与 IEMOCAP 相当（$\kappa = 0.27$，针对分类标签；Busso 等人，2008年），并反映了临床言语中情绪标注的固有难度（Kreiman 等人，1993）。我们注意到，对照数据仅有四名标注员（排除后为三名），这可能会影响其与五名标注员患者标注的可比性。

4.6 最终标签判定

算法1分层标签判定

要求： L：来自 n名标注员的标签列表，C：置信度分数
确保： $GoldLabel$：最终分配的标签
如果任一标签的多数计数 $x \in L \geq \lceil n/2 \rceil$ **则**
  $GoldLabel \leftarrow x$
否则如果中性与情绪之间存在并列 e**则**
  $GoldLabel \leftarrow e${非中性优先}
否则
  $W_i \leftarrow$ 计算每个并列标签的加权和 i 使用 C
  如果max(W) 是唯一的**则**
    $GoldLabel \leftarrow$ 最大值索引(W)
  else
    $GoldLabel \leftarrow$ NaN {标记为模糊}
  结束如果
结束如果

为确定每个话语的金标准标签，我们按照算法 1 中所述实施了分层仲裁流程。该流程优先考虑共识和情绪敏感性，并利用置信度分数来解决平局。

对于阿尔茨海默病患者数据，五位标注员的标签均被采用，多数阈值为 $\geq$3。对于对照数据，由于 L3 的标注模式与上述模式不同，因此排除了其标注，并对其余三位标注员应用相同的算法，多数阈值为 $\geq$2。

导致 *NaN* 的话语（阿尔茨海默病组 137 条，对照组 9 条）被归类为模糊，并在第5.2节中单独分析。

5 语料分析

5.1 情绪分布

表 5和图 1展示了带有标签的 PAR 话语的情绪分布。与对照组（5.7%；$\chi^2(1) = 38.45$，$p < .001$）相比，阿尔茨海默病患者表现出显著更高的非中性情绪比例（16.9%）。

喜悦是两组中最常见的非中性情绪（阿尔茨海默病组：7.6%，对照组：3.3%），其次是惊讶（阿尔茨海默病组：4.2%，对照组：0.8%）。悲伤、愤怒、厌恶和恐惧总体上都很少见，但在阿尔茨海默病语音中出现得更频繁。

情绪	AD n	AD %	对照组 n	对照组 %
中性	511	83.1	689	94.3
Joy	47	7.6	24	3.3
惊讶	26	4.2	6	0.8
悲伤	15	2.4	5	0.7
愤怒	8	1.3	1	0.1
厌恶	6	1.0	6	0.8
Fear	2	0.3	0	0.0
非中性	104	16.9	42	5.7
总计	615	100	731	100

表5：已标注PAR话语的情绪分布。

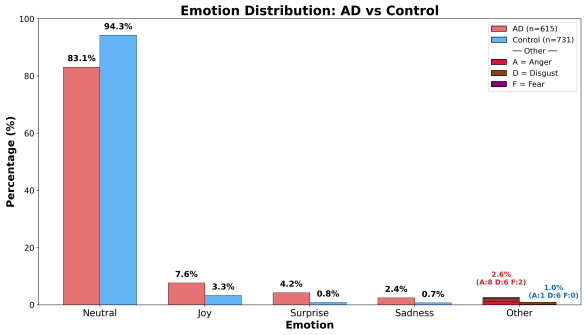


图1：阿尔茨海默病患者（n=615）与健康对照组（n=731）的情绪分布对比。罕见情绪（愤怒、厌恶、恐惧）合并为“其他”，并显示其计数。阿尔茨海默病患者的非中性情绪比例显著较高（16.9% vs. 5.7%；$\chi^2 = 38.45, p < .001$）。

5.2 模糊案例

在标记为模糊的146条话语中，137条出现在阿尔茨海默病语音中，9条出现在对照组语音中。对比模糊与非模糊的阿尔茨海默病话语后发现，模糊案例的基频、基频方差、响度和谐噪比显著较低，词语也更少——这表明模糊往往源于韵律线索不足。

分析揭示了两种模糊模式：（1）大多数案例呈现平坦或减弱的声学特征，无法提供足够的韵律线索来识别情绪；（2）少数案例的基频方差较高，可能与多种情绪类别相符。表 6 展示了后一种模式。

这些发现表明，AD情绪标注中的歧义有两个来源：情感平淡导致韵律信息不足，以及韵律线索确实可以支持多种情感解释。

话语	Spk	基频（半音）		基频方差		响度	
		eGe	spk	eGe	z	eGe	z
“她不知道该怎么想这件事，真见鬼。”	010（男性，69岁）	21.0	−0.1	0.28	+1.04	0.17	−0.47
“其中一个孩子要挨一下了”	018（男性，66岁）	31.4	−4.2	0.29	+0.95	0.07	−0.58

标签：（上）厌恶(2)、愤怒(2)、悲伤(1)；（下）恐惧、惊讶、中性、喜悦

表6：具有代表性的模糊话语（基频方差升高型）。eGe = eGeMAPS值；spk/z =说话人归一化。

5.3 定性分析

对非中性话语的考察揭示了与我们在校准工作坊中确立的标注标准一致的话语模式（第4.4节）。

作为面子维护的笑声。许多“喜悦”标签（47条阿尔茨海默病话语）与笑声标记同时出现（例如，“&=笑”）。如第4.4节所述，笑声子类型依据韵律线索加以区分：带有升调的愉快笑声被标记为喜悦，而带有降调的无助笑声则被标记为悲伤。这一区分至关重要，因为阿尔茨海默病语音中的笑声往往在交流困难时发挥保全面子的功能，而非表达真正的快乐。

惊讶与察觉。惊讶标签（26条阿尔茨海默病话语）常出现在参与者注意到图片中的混乱元素时，通常伴有感叹词“哦”和音高变化。

悲伤与共情。悲伤标签（15条AD话语）的特征为降调与较慢语速，常与对所描绘人物的同情表达同时出现。这些话语表明对叙事内容的情感投入得以保留。

愤怒与挫折。愤怒标签（8条AD话语）反映了与任务相关的挫折，例如“到底还有什么？”和“我不知道！”。此类表达可能表明其意识到沟通困难。

临床意义。这些模式表明，阿尔茨海默病语音的情绪标注需要关注语用与互动功能，而不仅仅是表层线索。以笑应对、挫折所体现的自我觉察以及对角色的共情，均反映出得以保留、且可能与语言障碍并存的社会情绪加工。

5.4 语言学与声学分析

表7展示了说话人层面的语言与声学特征。阿尔茨海默病患者产出的话语略短（平均话语长度 = 7.84个词语），相比对照组（平均话语长度 = 8.70; $t = -2.13$, $p = .035$, $d = -0.41$）。每次会话的总词语数与形符类符比（词汇多样性比率）在组别之间无显著差异。

声学特征使用openSMILE提取（Eyben 等人，2010年），采用eGeMAPS特征集（Eyben 等人，2016年），其中包括韵律特征（基频、响度）、嗓音质量指标（基频微扰、微颤、谐噪比）以及频谱参数。值得注意的是，H1-H2（第一与第二谐波之间的振幅差）用于指示发声类型：较高值表示更明显的气声嗓音质量，而较低值则提示挤压发声或嘎裂发声（Gordon与Ladefoged，2001年；Hanson，1997年；Keating等人，2015）。

与对照组相比，阿尔茨海默病患者的谐噪比显著较高（$d = 0.42$，$p = .033$），但这种差异可能反映的是录音条件的变化，而非嗓音质量的差异。其他声学特征（基频、响度、微颤、H1-H2）在说话人层面上均未显示组间显著差异。大多数声学特征不存在组间差异，这与使用较小对照样本的既往研究形成对比；我们的匹配队列每组包含54名说话人，提供了更均衡的比较。

5.5 情绪与声学

基频归一化。基频以半音为单位进行转换，并以每位说话人的平均基频作为参照（Jeong 和 Wedel，2026；Rose，2002；Nolan，2009）：

$$\mathrm{F0}_{\mathrm{st}} = 12 \times \log_2\left(\frac{\mathrm{F0}_{\mathrm{Hz}}}{\mathrm{F0}_{\mathrm{mean,speaker}}}\right) \qquad (1)$$

这种说话人相对归一化将语调动态与生理基线区分开来，从而能够在说话人之间进行比较。其他声学特征则在说话人内部进行 Z 分数标准化。

在阿尔茨海默病患者中，我们考察了声学特征是否能够区分情绪类别

	语言学（n=54 阿尔茨海默病，54 对照组）			声学（n=54阿尔茨海默病，54 对照组）				
组别	MLU	词语	TTR	响度	微扰	HNR	F0	H1-H2
AD	7.84 (2.17)	106.9 (67.4)	0.58 (0.10)	0.41 (0.37)	1.40 (0.26)	4.89 (2.29)	28.63 (3.89)	2.69 (3.90)
对照组	8.70 (2.06)	118.3 (51.8)	0.59 (0.07)	0.50 (0.45)	1.45 (0.28)	3.96 (2.19)	27.16 (4.99)	3.39 (3.46)
p	.035*	.328	.422	.291	.296	.033*	.091	.326
d	−.41	−.19	−.16	−.20	−.20	.42	.33	−.19

表7：语言学和声学特征（说话人层面）。数值为均值（标准差）。微颤（微扰）和谐噪比的单位为分贝，基频的单位为半音，响度（响度）的单位为宋。*$p < .05$，**$p < .01$，***$p < .001$。

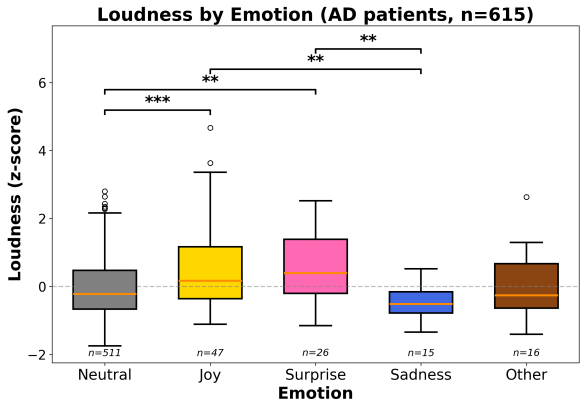


图2：阿尔茨海默病患者按情绪类别划分的说话人标准化响度。数量较少的情绪（愤怒、厌恶、恐惧）合并为“其他”。星号表示成对比较存在显著差异（Tukey诚实显著差异检验）：**$p < .01$，***$p < .001$。其他声学特征（基频、基频微扰、微颤、谐噪比）均未显示显著差异（见附录C）。

使用这些说话人标准化数值。罕见情绪（愤怒、厌恶、恐惧；每种 $n < 10$）被合并为“其他”类别（n=16），以提高统计效力。单因素方差分析显示，情绪对响度有显著影响（$F(4, 610) = 8.48$，$p < .001$，$\eta^2 = 0.053$），但对基频没有显著影响（$F(4, 610) = 2.11$，$p = .078$）。事后Tukey诚实显著差异检验显示，喜悦和惊讶话语的响度显著高于中性和悲伤：喜悦 vs. 中性（$\Delta = +0.59$，$p < .001$），惊讶 vs. 中性（$\Delta = +0.64$，$p = .004$），喜悦 vs. 悲伤（$\Delta = +1.00$，$p = .002$），以及惊讶 vs. 悲伤（$\Delta = +1.04$，$p = .003$）。图 2 展示了这些模式。说话人归一化后，其他声学特征（基频、基频微扰、微颤、谐噪比）在不同情绪之间均未显示出显著差异。

组别 × 情绪交互作用。 为考察阿尔茨海默病患者与对照组说话人在声学实现情绪的方式上是否存在差异，我们拟合了以说话人为随机截距的线性混合效应模型：特征 ~ 组别 × 情绪 +(1|说话人)。基频的组别 × 悲伤交互作用达到显著水平（$\beta = -5.20$，标准误 = 2.28，$p = .023$）：在表达悲伤时，对照组说话人的基频相较其中性基线显著下降（$\Delta = -3.45$ 个半音），而阿尔茨海默病患者的基频几乎没有调制（$\Delta = +0.11$ 个半音）。在说话人层面进行自助重抽样（1,000 次迭代）得到该交互作用效应的 95% 置信区间为 [−5.70，−0.22]，进一步确认了统计显著性。微颤也观察到类似模式（$\beta = -0.53$，p= .017）。这些发现表明，尽管阿尔茨海默病患者表达的情绪内容更多（16.9% 对 5.7% 的非中性表达），但其对这些情绪的韵律实现有所减弱——我们将这一模式称为声学平坦化。

6 讨论

6.1 解读阿尔茨海默病中的更高情绪表达

我们的研究发现，阿尔茨海默病患者表达的非中性情绪（16.9%）显著多于健康对照组（5.7%；$\chi^2 = 38.45$，$p < .001$），这一结果值得谨慎解读。乍看之下，这种模式似乎有悖直觉：人们可能会预期，认知能力下降会表现为情感淡漠或情绪表达能力减弱。然而，我们的分析表明，阿尔茨海默病语音中的情绪表达反映了仍然保留的社会情绪加工能力；这种能力与语言障碍并存，有时还能够对其进行补偿。

如第4.4节所述，阿尔茨海默病语音中的笑声往往具有保全面子的功能，而非表达真正的快乐。这一发现对情绪识别系统具有重要意义：在健康人语音上训练的简单分类器可能会将笑声标记直接映射为喜悦，从而系统性地误判这些补偿性实例。因此，阿尔茨海默病语音中较高的非中性表达语速，可能反映了在语言困难时刻更多地运用情绪表达作为

应对策略（Glenn，2003年）。

6.2 Cookie Theft任务作为情绪诱发因素

Cookie Theft 图片描述任务虽然主要被设计为一种认知—语言学评估，但它创造了可能自然诱发情绪反应的条件——尤其是对于经历找词困难的个体。该任务将说话人置于必须展示能力的公开表现情境中，而所描绘的混乱场景为情感投入提供了充足的材料。

我们的分析表明，阿尔茨海默病患者对其环境仍保持情绪反应：惊讶标志着注意到意外元素的时刻，而悲伤则反映了对图中人物的共情。这些能力对于理解阿尔茨海默病中的生活质量和社会功能可能具有临床意义。

6.3 阿尔茨海默病语音中情绪的声学特征

在阿尔茨海默病患者中，响度显著区分了情绪类别，喜悦和惊讶话语的响度高于中性话语。这些模式与在健康人群中记录的情绪的声学相关特征大体一致，表明情绪与声学表达之间的映射在阿尔茨海默病中至少部分得以保留。

一项比较阿尔茨海默病组与对照组的探索性分析揭示了一种值得关注的模式。尽管阿尔茨海默病患者表达非中性情绪的比例更高（16.9% 对 5.7% ），但不同情绪类别之间的韵律差异较小。表达悲伤时，对照组说话人的基频调制幅度较大（相较于中性基线为 $\Delta = -3.45$ 个半音），而阿尔茨海默病说话人的基频变化极小（$\Delta = +0.11$ 个半音）。组别 $\times$ 情绪交互作用具有统计显著性（$\beta = -5.20$，$p = .023$），并经自助法验证（95% 置信区间：$[-5.70, -0.22]$）。不过，鉴于非中性情绪的样本量较小，尤其是悲伤情绪（对照组中 3 名说话人的 n=5 条话语，阿尔茨海默病组中 11 名说话人的 n=15 条话语），应谨慎解读这一发现。这一初步模式——我们暂且称之为声学平坦化——或许表明，阿尔茨海默病患者虽然能够感知或意图表达情绪，但其韵律实现有所减弱。

如果在更大样本中得到重复验证，这一发现将与阿尔茨海默病中已有报告的韵律障碍相符（Themistocleous等人，2020），并将对情绪识别系统产生影响：处理阿尔茨海默病语音时，系统可能需要依赖多模态线索，而非仅依靠韵律。

尽管存在声学平坦化，专家标注员仍能识别出阿尔茨海默病语音中显著更多的非中性情绪，这一发现或许有助于解释自动化方法得出的结果为何存在差异。 Chou 等人(2025年通过仅基于声学特征的机器学习，发现阿尔茨海默病组和对照组之间与情绪相关的差异极小。自动化系统主要依赖在阿尔茨海默病中有所减弱的声学特征，而人工标注员则会综合韵律、语言学和语用线索，例如共情的词汇标记（“可怜的母亲”），或在遇到词语提取困难后为保全面子而发出的笑声。这表明，临床言语中的情绪识别可能需要超越单纯声学分析的多模态方法，也凸显了专家标注语料库对于训练此类系统的价值。

经过说话人归一化后，嗓音质量指标（基频微扰、微颤、谐噪比）未能显著区分不同情绪。这一无显著差异的结果可能反映了疾病相关的基线差异，也可能是非中性情绪的样本量不足所致。

6.4 临床语音情绪标注的挑战

我们的评分者间信度（Fleiss’ $\kappa = 0.23–0.31$）按传统标准属于中等，但与其他自发语音情绪语料库的结果相当（Busso 等人，2008年）。造成这一难题的因素有几个：（1）自发言语中的情绪本质上具有模糊性，反映了情绪混杂以及感受与表达之间的不一致；（2）即使患者意图表达中性情绪，阿尔茨海默病语音中的非典型韵律也可能被感知为带有情绪；（3）临床群体中情绪表达的语用功能有所不同，这一点已在我们的校准工作坊中得到记录（第4.4节）。

这些挑战凸显了不仅公开我们的标注，也公开标注指南和校准工作坊材料的重要性。

6.5 语音质量测量的无显著结果

嗓音质量指标（基频微扰、微颤、谐噪比）在说话人归一化后，未能显著区分情绪

。尽管谐噪比在说话人层面显示出显著的组别差异（表7），但这并未转化为说话人内部的情绪层面区分。

我们还考察了 H1-H2，它用于衡量发声类型（例如，气声与挤压声（更多信息请参见 Keating 等人 (2015 ））），但与整段话语相比，从元音的稳态部分提取这一指标最为可靠。我们在话语层面进行的提取会对辅音、停顿和不流利现象取平均，因此可能引入大量噪声。v2.0 中的音素层面分段将有助于更精确地测量发声相关特征。

零结果的一种可能解释是，嗓音质量反映的是稳定的疾病相关嗓音变化，而非短暂的情绪状态。阿尔茨海默病相关的神经退行性病变可能影响喉部运动控制（Themistocleous等人，2020），从而造成基线嗓音质量差异，并使其在不同情绪状态下持续存在。另一种可能是，非中性情绪的样本量不足以检测出嗓音质量差异。未来若扩大非中性情绪样本，并开展音素层面的声学分析，或可阐明嗓音质量是否有助于区分阿尔茨海默病语音中的情绪。

7 结论

我们介绍了 DementiaBank-Emotion v1.0，这是首个针对阿尔茨海默病语音的多标注者情绪标注语料库。我们的主要发现是，阿尔茨海默病患者表达的非中性情绪显著多于对照组（16.9% 对 5.7%； p<.001 ）。探索性声学分析表明，两组可能存在分离现象：对照组说话人在表达悲伤时，基频调制幅度较大（$\Delta = -3.45$ 半音），而阿尔茨海默病说话人的变化极小（$\Delta = +0.11$ 半音；交互作用 p =.023）。然而，这种声学平坦化模式基于有限样本（悲伤： n=5名对照组说话人，n=15名阿尔茨海默病说话人），在重复验证之前应视为初步结果。在阿尔茨海默病语音中，响度能够区分不同情绪类别（喜悦和惊讶时更高），表明情绪—韵律映射部分保留。

我们发布该语料库、标注指南和校准研讨会材料，以支持未来的研究。计划中的扩展（v2.0）将纳入纵向数据，并考察情绪表达与认知严重程度之间的关系。此外，正在进行的音素层面由受训语音学家进行的重新对齐将支持细粒度的声学语音分析，包括发声类型、元音空间动态变化和共振峰轨迹。这种音段分析还扩展到填充停顿和话语标记语：阿尔茨海默病患者使用*uh* 更多，但使用*um*少于对照组（Yuan等人，2020），而韵律则能消解话语标记语的功能歧义（Jeong和Park，2017年）。在主观化框架（Traugott，1995；Schiffrin，1987）中，这类标记不仅指示认知加工，还指示情感立场；当词汇资源受限时，它们将细粒度的语音分析与情绪表达的语用学联系起来。

局限性

有几项局限性需要说明。

对照数据校准研讨会。阿尔茨海默病数据和对照数据均来自 ADReSS 2020 挑战赛数据集，且人口统计学特征匹配（每组 54 名说话人，年龄和性别均衡），但由于时间限制，阿尔茨海默病数据经过了三轮校准研讨会，对照数据则没有。这导致阿尔茨海默病数据的歧义率（18.2%）高于对照数据（1.2%），因为标注员通过研讨会讨论，为阿尔茨海默病数据形成了共同的隐性标准。未来版本应纳入针对对照组标注的校准研讨会，以确保流程一致性。

类别不平衡。数据集中的话语严重偏向中性（阿尔茨海默病组占 >83%，对照组占 >94%），而罕见情绪（恐惧、厌恶、愤怒）各自出现的次数不足 10 次。这种类别不平衡限制了我们描述罕见情绪声学特征的能力，也给训练监督分类器带来了挑战。正如一位审稿人所指出的，未来进行声学分析时，可能需要将罕见情绪合并为更宽泛的类别（例如“负效价”），以提高统计功效。

任务情境。 Cookie Theft任务代表单一交际情境——在评估情境中诱发的图片描述。情绪表达模式在会话、叙事或现实情境中可能存在显著差异。我们关于保全面子的笑声以及得以保留的情绪—韵律映射的发现，需要在多样化交际任务中加以复现，方可进行推广。

声学分析粒度。 声学特征在话语级提取，使用原始CHAT转录时间戳；这适用于整体韵律测量指标（基频、响度），但无法对话语内部动态进行细粒度分析。曾尝试强制对齐，但由于阿尔茨海默病语音特有的不流畅现象，失败率约为20%。版本 2.0将纳入经人工校正的音段层面切分。

感知情绪与体验情绪。 我们无法获取真实情绪状态。我们的标注反映的是根据声学和语言线索感知到的情绪，这些情绪未必与说话人实际体验到的情绪相符。这一局限是大多数情绪标注工作的固有特点，但在临床人群中尤为突出，因为情绪的表达成分和体验成分可能彼此分离。

评分者间信度。 我们的 Fleiss’ κ 值（0.23–0.31）虽不高，但与 IEMOCAP（$\kappa = 0.27$；Busso 等人，2008年）相当。中等程度的一致性反映了临床语音情绪中真实存在的歧义，而非标注噪声；不过，使用该语料库的用户应在下游应用中考虑标签不确定性。

伦理考量

数据来源与知情同意。 本研究使用的语音数据来自 DementiaBank Pitt 语料库（Becker 等人，1994年），该语料库是在匹兹堡大学获得机构审查委员会批准后收集的。所有参与者在原始数据收集时均提供了知情同意。我们的标注工作仅为现有公开数据添加元数据，不涉及新的人体受试者研究；尽管如此，我们仍获得了加利福尼亚大学欧文分校机构审查委员会的豁免（IRB # 3795）。

隐私与可识别性。 尽管 DementiaBank 语料库可公开用于研究，但语音录音本质上具有可识别性。除 DementiaBank 中已有的信息外，我们不会发布任何其他可识别信息。本语料库中的说话人 ID 与 ADReSS 挑战赛的标识符相对应，这些标识符已进行假名化处理。

弱势群体。阿尔茨海默病患者构成了一个弱势群体，其持续同意能力有所减弱。我们承认，为无法审阅或质疑我们所给标签的人标注情绪状态，存在伦理上的复杂性。我们的标注反映的是根据外在线索感知到的情绪，不应被解读为对参与者内在状态的断言，也不应被用于对个体患者作出临床判断。

滥用风险。 情绪识别技术引发了人们对监控和操纵的担忧。虽然本语料库旨在用于临床沟通和辅助技术开发方面的研究，但我们也承认情绪识别系统可能遭到滥用。我们鼓励用户考虑下游应用的伦理影响，并优先选择有益于痴呆患者及其照护者的应用。

双重用途与人工智能辅助。 在准备稿件时，我们使用了 Claude（Anthropic）进行语法检查并获得编辑辅助。科学内容、分析和解读均由人类作者独立负责。

数据可用性

DementiaBank-Emotion语料库将在论文发表后通过 DementiaBank（https://dementia.talkbank.org/）发布。

致谢

我们感谢 UC Noyce倡议提供的支持。我们感谢三位匿名审稿人提出的建设性意见，这些意见大幅改进了本文。我们使用 Claude（Anthropic）进行语法检查和编辑辅助。

参考文献

Samrah Ahmed、Anne-Marie F Haigh、Celeste A deJager 和 PeterGarrard。2013年。连贯语音作为尸检证实的阿尔茨海默病疾病进展的标志。《*Brain*》，136(12)：3727–3737。

James T Becker、François Boller、Oscar L Lopez、Judith Saxton 和 Karen L McGonigle。1994年。阿尔茨海默病的自然病程：研究队列描述及诊断准确性。《神经病学档案》，51(6)：585–594。

Dwight L Bolinger. 1958. 英语音高重音理论。*Word*, 14(2-3)：109–149。

Carlos Busso、Murtaza Bulut、Chi-Chun Lee、Abe Kazemzadeh、Emily Mower、Samuel Kim、Jean- nette N Chang、Sungbok Lee 和 Shrikanth S Narayanan。2008年。IEMOCAP：交互式情绪双人动作捕捉数据库。语言资源与评估，42(4)：335–359。

Carlos Busso、Srinivas Parthasarathy、Alec Burmania、Mohammed AbdelWahab、Najmeh Sadoughi 和 Emily Mower Provost。2017年。MSP-IMPROV：用于研究情绪感知的双人互动表演语料库。*IEEE Transactions on Affective Computing*，8(1):119–130。

Shih-Han Chou、Miini Teng、Harshinee Sriram、Chuyuan Li、Giuseppe Carenini、Cristina Conati、Thalia S Field、Hyeju Jang 和 Gabriel Murray。2025年。结合面部和眼动追踪数据对阿尔茨海默病进行多模态分类。发表于*MachineLearningforHealth(ML4H)*，第219–232页。PMLR。

Paul Ekman。1992年。关于基本情绪的论证。*Cognition& Emotion*，6(3-4):169–200。

Florian Eyben、Klaus R Scherer、Björn W Schuller、Johan Sundberg、Elisabeth André、Carlos Busso、Laurence Devillers、Julien Epps、Petri Laukka、Shrikanth S Narayanan及其他1位作者。2016年。用于语音研究和情感计算的日内瓦极简声学参数集 (GeMAPS)。*IEEETransactions on Affective Computing*，7(2):190–202。

Florian Eyben、Martin Wöllmer与Björn Schuller。2010年。openSMILE：慕尼黑多功能快速开源音频特征提取器。载于第*18*届*ACM*国际多媒体会议论文集，第1459–1462页。Kathleen C Fraser、Jed A Meltzer与Frank Rudzicz。2016年。语言特征识别叙事语音中的阿尔茨海默病。*JournalofAlzheimer'sDisease*，49(2):407–422。Phillip Glenn。2003年。*Laughter inInteraction*。剑桥大学出版社，剑桥。Matthew Gordon与Peter Ladefoged。2001年。发声类型：跨语言综述。*JournalofPhonetics*，29(4):383–406。Helen M Hanson。1997年。女性说话人的声门特征：声学相关因素。*TheJournal oftheAcousticalSociety of America*，101(1):466–481。Julie D Henry、Peter G Rendell、Andrew Scicluna、Maree Jackson与Louise H Phillips。2009。阿尔茨海默病中的情绪体验、表达与调节。*Psychologyand Aging*，24(1):252–257。Cheonkam Jeong与Seongjin Park。2017年。韵律在话语中的作用：韩语*com*案例研究。载于*SICSS 2017*论文集，第1–2页，首尔，韩国。首尔国立大学。

Cheonkam Jeong与Andrew Wedel。2026。对比特定最小对立组竞争项对声调发生音变中韩语词首塞音起声时间与基频语音线索过度发音的影响。*Journal ofPhonetics*，115:101474。

Patricia Keating、Marc Garellek 和 Jody Kreiman。2015。不同类型嘎裂嗓音的声学特性。载于第十八届国际语音科学大会论文集，第 1–5 页，格拉斯哥。

Jody Kreiman、Bruce R Gerratt、Gail B Kempster、An- drew Erman 和 GeraldS Berke。1993。嗓音质量的感知评估：综述、教程及未来研究框架。《言语、语言与听力研究杂志》，36(1):21–40。

Alyssa M Lanzi、Anna K Saylor、Davida Fromm、Hou- jun Liu、Brian MacWhinney 和 Matthew L Cohen。2023。DementiaBank：理论依据、方案及示例性分析。《美国言语语言病理学杂志》，32(2):426–438。

Steven R Livingstone 和 Frank A Russo。2018。瑞尔森情绪语音与歌曲视听数据库（RAVDESS）：一组北美英语中动态、多模态的面部和声音表达。*PLOS ONE*，13(5):e0196391。

Saturnino Luz、Fasih Haider、Sofia de la Fuente、Davida Fromm和Brian MacWhinney。2020。通过自发言语识别阿尔茨海默病性痴呆：ADReSS 挑战赛。载于*INTERSPEECH 2020*会议论文集，第2172–2176页。国际语音通信协会。Kimberly D Mueller、Bruce Hermann、Jonilda Mecollari和Lyn S Turkstra。2018。轻度认知障碍与阿尔茨海默病中的连贯语篇和语言：图片描述任务综述。《临床与实验神经心理学杂志》，40(9):917–939。Francis Nolan。2009。说话人识别的语音学基础。载于William J Hardcastle、John Laver和Fiona E Gibbon编，《语音科学手册》，第2版，第700–724页。Wiley-Blackwell，牛津。Philip Rose。2002。《法庭说话人识别》。Taylor & Francis，伦敦。Deborah Schiffrin。1987。《话语标记语》。剑桥大学出版社，剑桥。Elisabeth Selkirk。1995。句子韵律：语调、重音与韵律分组。载于John A. Goldsmith编，《音系理论手册》，第550–569页。Blackwell，马萨诸塞州剑桥。Christina Sobin和Murray Alpert。1999。言语中的情绪：恐惧、愤怒、悲伤和喜悦的声学属性。《心理语言学研究杂志》，28(4):347–365。

Charalambos Themistocleous、Marie Eckerström和Dimitrios Kokkinakis。2020。嗓音质量与言语流利度区分轻度认知障碍个体与健康对照组。*PLOS ONE*，15(7):e0236009。

Elizabeth Closs Traugott. 1995. 语法化中的主观化。载于 Susan Wright 和 Dieter Stein 编，《主观性与主观化》，第 31–54 页。剑桥大学出版社，剑桥。

Jiahong Yuan、Yuchen Bian、Xingyu Cai、Jiaji Huang、Zheng Ye 和 Kenneth Church。2020。用于检测阿尔茨海默病的不流畅现象与预训练语言模型微调。载于*INTERSPEECH 2020* 论文集，第 2162–2166 页。

A 标注指南

提供给评定者的完整标注指南收录于以下页面。这些指南详细说明了用于区分临床言语中Ekman的六种基本情绪与中性状态的声学和语言学标准。

标注指南（第 2.0 版，更新于2024年6月9日）

1. 引言

- 目标
  - 目标 3 的一项任务是对话语的情绪进行标注。本文档将介绍对每条话语进行标注时应采用的方法。
- 任务
  - 你的任务是聆听临床医生（INV）与患者（PAR）之间的对话，仔细阅读对应文本，为其添加标签，并说明你对该标签的置信度。有关流程，请参见 4.2. 流程。
  - 我们使用 7 种情绪类别（艾克曼六类情绪 + 中性），其定义见 5.1. 七种情绪类别。
- 待标注的数据
  - 共 495 条话语（包括表头）。说明请参见 4.1. 材料。
  - 在以你姓名命名的文件夹中
    - 共享给你的文件夹中有 1 个 Google表格文本文件
    - 同一文件夹中有 5 个音频文件
- 时间安排
  - 请先做“第一件事”，然后我们再谈
- 文件位置
  - 第一 第一
  - 第二 第二
  - 第三 第三

2. 关键术语

2.1. 话语

就本任务而言，我们将术语**话语**定义为以10毫秒切分的单个单元。在某些情况下，这对应于一个没有停顿的单个句子；而在其他情况下，它实际上可能由不止一个句子组成。偶尔，单个句子甚至会被拆分为两个话语。

2.2. 情绪

在本任务中，情绪指说话人在一段话语中表现出的离散情绪。我们使用7个类别（艾克曼六类情绪+ 中性）。情绪从下文所述的标签集合中选取。

3. 数据描述

在此任务中，我们使用来自*DementiaBank*英语皮特语料库的开放数据集（Becker 等人，1994年）。该语料库提供了临床医生与患者之间的对话。痴呆症患者需要描述下方的“偷饼干”图片。

The "cookie theft" picture from the Boston Diagnostic Aphasia Examination



No

示例对话的部分内容如下。可以看到，临床医生（INV）请患者（PAR）描述“偷饼干”图片，患者则描述自己看到的内容。

说话人	**话语**
INV	你在那张图片里看到什么情况?
PAR	嗯。
INV	发生什么事了?
PAR	嗯，孩子正在偷饼干罐里的饼干。
PAR	而母亲正在洗餐具，却没注意到这一点
PAR	它正在地板上流淌。
INV	嗯。
PAR	而他的凳子就快把他摔下来了&=笑

4. 标注流程

4.1. 材料

● 1 个文本文件：the_second_labeling.xlsx ●

同一文件夹中的 7 个音频文件

○ 除第一个音频文件（第三_s108_122.wav）因包含大量行外，每个音频文件都包含特定的行和音频。○ 例如，文件**第三_s114_201.wav**涵盖 S110 与 S114 之间的对话，话语范围从第 123 行到第 201 行。

文件名	**行**	**音频**
第三_s108_122.wav	第003–122行	S108
第三_s114_201.wav	第123–201行	S110-114
第三_s125_306.wav	第202–306行	S116-125
第三_s130_424.wav	第307至424行	S126-132
第三_s139_504.wav	第425至504行	S135-139
第三_s145_597.wav	第505至686行	S140-145
第三_s156_686.wav	第598–686行	S148–156

4.2. 流程

1. 执行此任务时，您应在聆听音频的同时打开电子表格。请选择安静的地方进行工作，并使用耳机，以确保能够清楚地听到音频。

2. 听完每条话语后，暂停录音，然后将情绪标签输入电子表格 D 列的“情绪类别”栏。此外，请在 E 列（“您有多确定？”）中标明您对该标签的置信度。随后，您可以继续播放录音，以查看下一条话语。3. 如果不确定，请听前一条和/或后一条话语。
4. 每条话语都应赋予一个单独的标签。该标签可以根据参与者说出的词语、其说话方式，或两者综合确定。

---

5. 情绪标签

5.1. 七种情绪类别

在进行标注时，我们使用由埃克曼的六种普遍情绪+ 和一个**中性**标签组成的集合。该标签集合如下：

1. **愤怒**：说话人感到愤怒、不安，并通过词语、语气或二者表现出来。 2. **厌恶**：说话人感到厌恶、不快、反感或憎恶。 3. **恐惧**：说话人害怕某件事。 4. **中性：**（没有明显情绪）——说话人既没有表现出任何情绪，也没有表达任何观点。他们可能是在陈述、描述或提供有关自己在图片中看到的内容的信息。 5. **喜悦**：说话人感到开心、玩得愉快，或以其他方式享受某件事。 6. **悲伤**： 传达出失落、无助或失望。 7. **惊讶**：发生了令人惊讶的事情，说话人突然获得了新的、出乎意料的信息。

5.2. 如何决定选择哪个情绪标签

为一段话语选择正确的标签通常很直接，但有时也会有些棘手。一般来说，请遵循以下指南：

●如果一段话语没有明显的情绪信息，就将其标记为**中性**。否则，请使用下面的具体标签。 ●如果一段话语大部分内容是中性的，但其中有一部分包含明显的情绪内容，则根据该非中性部分表达的情绪进行标记。●如果一段话语包含两种情绪，请按以下方式处理：○ 如果其中一种情绪明显强于另一种，就选择较强的情绪

○ 如果一种情绪在话语中占主导，请选择主导情绪**注意！！！**

● 在个人印象和特定词语的使用方面应谨慎。例如，避免仅根据“刻薄”之类的词语进行标注，因为这些词语可能无法准确反映患者情绪。此类词语的内涵以及文化因素都可能影响标注。● 了解患者的整体言语风格会有所帮助。请记住，言语的表达方式比所用的特定词语更为重要。应关注患者所表达的情绪，而不是推测这些情绪的成因。

具有言语特征的具体标签：

1. **喜悦** a. “笑声”是一个强烈的指标，但应谨慎标注，因为它可能与讽刺重叠。b. 患者感到喜悦时，其言语特征表现为高音量、中等音量变化幅度和/或高音高变化幅度。c. 他们的语速适中，言语和停顿的持续时间都较长。d. 临床医生鼓励患者说话时，患者的言语通常会变得夸张，音高和强度会升高或上扬。其言语风格通常带有幽默语气。

2. **悲伤**

   a. 如果患者说话含糊和/或不清楚、模棱两可，则使用“悲伤”，但应谨慎标注，因为它可能与恐惧重叠。b. 患者感到悲伤时，其言语特征表现为低音量、高音量变化幅度和/或低音高变化幅度。其整体音高较低。c. 他们的语速较慢，言语和停顿的持续时间都较长。可以发现许多停顿。

3. **恐惧**

   a. 当患者表达对某事的恐惧时，使用“恐惧”。恐惧的诱因包括周围环境的可见度降低、社交互动和/或拒绝、死亡与濒死等。b. 当患者感到恐惧时，其言语特征表现为高音量、低音量变化幅度和/或高音高变化幅度。其整体音高较高。c. 其语速很快，言语和

   p停顿的持续时间都很短。停顿不多。

4. **厌恶**

   a. 当患者表达对某事的轻蔑时，使用“厌恶”，但需谨慎标注，因为它可能与愤怒重叠。

5. **愤怒**

   a. 当患者感到愤怒时，其言语特征表现为高音量、高音量变化幅度和/或低音高变化幅度。其整体音高低。b. 其语速快，言语与停顿的时长均较短。停顿很少。

6. 惊讶

a. 当出现意外情况时，患者的音高会发生剧烈变化，但需谨慎标注，因其可能与喜悦重叠。b. 喘气也可作为一项指标。

7. 中性

a. 未观察到任何特定特征时，标注为“中性”。b. 患者逐步解释和/或描述某件事时，其言语往往平淡，语调单调。c. 使用描述性词语，而非表达情绪的词语，且不带有特定观点。

6. 参考文献

J. T. Becker、F. Boller、O. L. Lopez、J. Saxton 和 K. L. McGonigle，“阿尔茨海默病的自然史”，《神经病学档案》，第 51 卷，第 6 期，第 585 页，1994年。Ekman, P.（1999）。基本情绪。认知与情绪手册，*98*（45-60），16。Sobin 和 Alpert, M.（1999）。言语中的情绪：恐惧、愤怒、悲伤和喜悦的声学属性。《心理语言学研究杂志》，*28*（4），347–365。https:// doi.org/10.1023/a:1023237014909

B 完整歧义话语

表8：ADReSS2020 挑战赛数据集中的所有模糊话语（$n = 146$），包含声学特征，按基频方差 z 分数降序排列。说话人 ID（S0xx）对应 ADReSS 标识符。eGe = eGeMAPS（基频（半音）以 27.5 Hz 为基准；基频方差 = 标准差归一化；响度以宋为单位）。spk/z = 说话人归一化（基频（半音）以说话人均值为基准；其他特征在说话人内部进行 z 评分）。

		基频（半音）		基频方差		响度	
Spk	**话语**	eGe	spk	eGe	z	eGe	z
S095	“xxx还有别的吗？”	23.6	+3.8	0.52	+2.79	0.09	+0.74
S093	“而且（.）不知为何她一定是为.....感到难过的”	34.0	+3.7	0.52	+2.33	0.03	-0.62
S126	“看起来 +/。”	29.7	+2.3	0.34	+1.83	0.13	+1.74
S126	“(.) 而且还有别的东西在那边。”	20.5	-6.9	0.33	+1.67	0.03	-1.05
S124	“水槽的水正开着（g）xxx。”	21.4	-1.7	0.31	+1.47	0.03	-0.83
S111	“他的 k[er]@u。”	33.9	-0.5	0.27	+1.41	0.15	-1.11
S086	“他正试图自杀（g）xxx。”	32.3	+7.1	0.32	+1.35	0.13	+2.37
S093	“但她本会抓住他并救了他.....”	30.6	+0.3	0.38	+1.34	0.06	+0.31
S082	“它从水槽溅了出来，但没有从（.）从 +.....”	27.9	+0.9	0.39	+1.32	0.07	-0.35
S072	“（..）她一点也没有留意她的孩子。”	24.6	+1.9	0.42	+1.30	0.53	-1.62
S082	“突然间有人把一个盘子翻了过来。”	30.5	+3.5	0.38	+1.29	0.09	-0.28
S090	“嗯，让我看看。”	32.9	+3.8	0.46	+1.28	0.06	-0.98
S107	“（.....）xxx。”	28.6	+0.6	0.13	+1.14	0.08	-0.43
S125	“哦。”	29.6	-4.0	0.15	+1.14	0.14	+2.20
S080	“+< 好的。”	34.6	+17.3	0.31	+1.11	0.37	+1.45
S089	“所以她会找到她的 .”	26.7	-2.4	0.29	+1.10	0.14	+1.10
S090	“看起来像 +.....”	35.4	+6.2	0.44	+1.08	0.07	-0.92
S095	“然后她甚至都没在看他们。”	16.0	-3.8	0.27	+1.05	0.02	-0.52
S107	“xxx。”	28.5	+0.5	0.12	+1.00	0.29	+0.91
S082	“而母亲没有看到它，因为她在里面 ...”	27.3	+0.3	0.35	+0.96	0.09	-0.28
S126	“(.)还有另一个女孩(.)看起来像 .”	25.9	-1.5	0.29	+0.93	0.04	-0.61
S126	“(.)一些小结之类的东西(g) .”	25.5	-1.9	0.29	+0.83	0.04	-0.57
S111	“她。”	33.6	-0.8	0.23	+0.82	1.05	+1.05
S093	“可能把她的盘子掉了 &=笑 .”	36.3	+6.0	0.31	+0.81	0.08	+1.04
S139	“(..) xxx .”	26.8	-1.6	0.29	+0.80	0.02	-1.23
S125	“(..) 差不多就是这样。”	35.6	+2.0	0.14	+0.80	0.02	-0.60
S003	“他正在拿一块饼干，而且他分享一块饼干.....”	31.9	-1.3	0.21	+0.76	0.55	-0.41
S126	“我看到一个女孩站在那儿，或者——什么的。”	27.3	-0.1	0.28	+0.74	0.05	-0.36

（续见下页）

（接上页）

Spk	话语	基频（半音）		基频方差		响度	
		eGe	spk	eGe	*z*	eGe	*z*
S111	“+< xxx 充足的 xxx rice &=清嗓子 出院 .”	34.8	+0.4	0.23	+0.73	0.72	+0.27
S126	“(loo)ks like somebody took some 铅笔或什么(...”	25.8	-1.7	0.28	+0.73	0.08	+0.44
S125	“母亲正在把水洒出来 .”	34.3	+0.7	0.13	+0.72	0.02	-0.46
S093	“我不敢相信她会因为 孩子（因）为.....”	24.9	-5.3	0.29	+0.71	0.03	-0.72
S093	“换句话说，它是 +.....”	25.8	-4.5	0.29	+0.71	0.03	-0.80
S124	“&=（叹气）窗户开着。”	22.4	-0.7	0.24	+0.71	0.05	-0.30
S082	“我太、太想从中得到太多了 out_它。”	25.5	-1.5	0.33	+0.70	0.04	-0.50
S082	“只是它确实没有把它弄干。”	26.7	-0.3	0.33	+0.70	0.05	-0.44
S108	“嗯哼。”	35.6	+3.7	0.26	+0.67	0.16	-1.11
S082	“+(.) 但什么也没得到 你会想要想要.....”	30.7	+3.7	0.32	+0.65	0.13	-0.08
S090	“哦 +.....”	24.9	-4.2	0.38	+0.58	0.17	-0.30
S079	“而且她是她有有 +/。”	24.8	+0.0	0.32	+0.57	0.37	-0.50
S082	“+，可以说是一幅弱图像。”	33.9	+6.9	0.31	+0.55	0.29	+0.63
S082	“+< 不过，你会说这是一次蚀刻 一点点 .”	27.4	+0.4	0.31	+0.52	0.15	+0.02
S095	“他的凳子要倒了。”	38.3	+18.4	0.19	+0.51	0.04	-0.10
S082	“+< 而且有一个孩子要 在.....上撞出一道裂痕”	32.7	+5.7	0.30	+0.44	0.15	+0.02
S082	“+<&=笑声 xxx !”	35.2	+8.2	0.30	+0.43	1.06	+4.13
S111	“这个是，哦，我可以直接 +.....”	32.2	-2.2	0.20	+0.43	0.43	-0.42
S125	“有个饼干罐。”	34.6	+1.0	0.12	+0.43	0.03	-0.44
S093	“她，她在想，如果她真的..... 看到他们，她就.....” ...”	34.9	+4.6	0.25	+0.41	0.05	+0.08
S126	“一些 xxx 之类的东西。”	28.3	+0.9	0.26	+0.40	0.03	-0.81
S090	“（.....）哦，我看不懂 +.....”	28.1	-1.0	0.35	+0.37	0.05	-1.06
S096	“不，我看不太清楚， .”	28.1	-3.9	0.26	+0.36	0.16	-0.89
S089	“然后xxx母亲把餐具洗干（g） 餐具。”	30.0	+0.8	0.24	+0.32	0.06	-0.50
S020	“从外面看出去会很困难 窗户，然.....”	23.9	+0.9	0.22	+0.29	0.11	-1.30
S124	“xxx 那边的母亲，她正在洗—— 餐具。”	21.6	-1.4	0.20	+0.26	0.02	-0.92
S111	“xxx。”	39.1	+4.7	0.19	+0.25	0.66	+0.12
S051	“还有什么？”	24.6	-0.6	0.33	+0.25	0.57	+0.36
S017	“对，就是这样。”	37.0	+4.9	0.24	+0.23	0.42	+1.00
S107	“boing@o .”	27.1	-0.9	0.08	+0.21	0.09	-0.34
S090	“+<xxx 是一个 +/。”	28.3	-0.8	0.34	+0.21	0.43	+1.49
S107	“嗯。”	39.7	+11.6	0.08	+0.20	0.11	-0.23
S111	“+< xxx。”	35.5	+1.1	0.19	+0.19	0.84	+0.56

（续见下页）

（续自上一页）

Spk	话语	基频（半音） eGe	spk	基频方差 eGe	*z*	响度 eGe	*z*
S084	“你_知道，我，我不好意思，但是 你_知道，我，我当时是 +.....”	30.1	+2.9	0.30	+0.19	0.16	-0.12
S095	“那个小女孩正求他 给她一个。”	36.7	+16.8	0.14	+0.13	0.04	-0.20
S107	“xxx。”	27.3	-0.8	0.07	+0.12	0.04	-0.64
S095	“+<我会告诉你 +是什么/。”	40.5	+20.6	0.13	+0.12	0.22	+3.38
S126	“在我看来好像是一样的，除了 上面那些东西。”	28.1	+0.7	0.24	+0.09	0.05	-0.45
S124	“也许风正吹进来。”	24.2	+1.2	0.18	+0.09	0.11	+1.19
S093	“他正把一块饼干递给 她。”	30.2	-0.0	0.21	+0.08	0.03	-0.83
S126	“(.) xxx。”	26.2	-1.3	0.24	+0.03	0.05	-0.34
S126	“(.) 某种 xxx 平底锅，或者某- 瘦（g）。”	23.9	-3.5	0.24	+0.02	0.06	-0.24
S093	“他可能会把所有饼干 罐子都带走.....”	40.3	+10.0	0.19	-0.03	0.13	+2.59
S082	“真是太 +了.....”	35.1	+8.1	0.25	-0.10	0.07	-0.33
S082	“有时候我我看得很清楚， 而有时候 ...”	27.8	+0.8	0.25	-0.11	0.08	-0.29
S108	“是的。”	35.6	+3.7	0.18	-0.13	0.47	-0.37
S082	“我想这张图片里 那位母亲 ...”	24.4	-2.6	0.25	-0.15	0.07	-0.36
S126	“(.) 那个女孩就在那里。”	25.8	-1.6	0.23	-0.18	0.06	-0.12
S082	“&=清嗓子 well &=清嗓子 the kids are ...”	27.8	+0.8	0.25	-0.19	0.04	-0.50
S093	“但我不认为她会让 他们继续下去，我.....”	27.9	-2.4	0.17	-0.19	0.04	-0.50
S086	“就这样吗？”	30.5	+5.3	0.13	-0.28	0.08	+0.73
S095	“两个杯子，还有一套餐具洗好了。”	41.9	+22.1	0.08	-0.29	0.04	-0.16
S095	“他们的妈妈正在洗餐具。”	37.3	+17.5	0.08	-0.29	0.05	+0.08
S082	“他们，他们要去拿一些 饼干，从厨 ...”	24.3	-2.7	0.23	-0.34	0.09	-0.25
S029	“她显然非常心不在焉，白日 做梦，以至于她.....”	28.6	-0.0	0.22	-0.34	0.16	-1.08
S093	“+, 三人行，不过厨房 会一团糟。”	27.3	-3.0	0.14	-0.37	0.04	-0.31
S096	“我看到一点点。”	33.4	+1.4	0.20	-0.38	0.39	+1.06
S124	“嗯哼。”	19.6	-3.4	0.14	-0.45	0.02	-1.10
S093	“她正在洗餐具。”	41.8	+11.5	0.13	-0.45	0.05	+0.11
S107	“(...) 毛巾。”	27.4	-0.6	0.04	-0.46	0.02	-0.79
S090	“&=唱歌。”	25.5	-3.6	0.26	-0.46	0.14	-0.49
S093	“嗯，孩子们正在往上爬 而他正要 ...”	36.2	+6.0	0.13	-0.48	0.05	+0.13
S118	“+<哦 (.) 我不知道。”	31.7	+1.4	0.18	-0.50	0.74	+1.53
S107	“窗户 窗户 +.....”	28.4	+0.4	0.04	-0.50	0.11	-0.19

（下页继续）

（接上一页）

Spk	话语	基频（半音） eGe	spk	基频方差 eGe	*z*	响度 eGe	*z*
S124	“凳子在那级台阶上 xxx。”	17.3	-5.7	0.13	-0.52	0.03	-0.70
S126	“(..) 我看到这个，他们都看起来，看起来差不多一样 ...”	27.6	+0.1	0.21	-0.53	0.05	-0.45
S095	“水正从水槽里溢出来 .”	13.4	-6.4	0.04	-0.56	0.02	-0.53
S093	“也许她确实转过身去看 他们，并且&=笑着，那.....”	31.4	+1.1	0.12	-0.58	0.06	+0.25
S051	“而且，往窗户外看时，为什么她让他.....”	26.9	+1.7	0.23	-0.58	0.83	+1.57
S084	“它做到了 +/? ”	34.2	+7.0	0.22	-0.62	0.38	+0.85
S107	“女孩在帮男孩拿饼干罐。”	27.3	-0.8	0.03	-0.63	0.07	-0.47
S126	“（.）我看不到 xxx。”	30.7	+3.2	0.20	-0.65	0.07	+0.17
S082	“然后孩子们就这样 +.....”	24.4	-2.6	0.20	-0.65	0.05	-0.42
S107	“（.）车库。”	27.5	-0.5	0.03	-0.70	0.03	-0.74
S126	“我不知道。”	28.9	+1.5	0.19	-0.77	0.10	+0.89
S125	“男孩正从（.）一 张凳子上摔下来。”	33.3	-0.3	0.06	-0.78	0.05	+0.09
S107	“xxx。”	28.0	-0.0	0.02	-0.81	0.05	-0.62
S107	“正在（g）洗餐具。”	26.5	-1.5	0.02	-0.81	0.05	-0.63
S095	“她想吃掉它。”	0.0	-19.9	0.00	-0.82	0.02	-0.51
S095	“她正指着自己的嘴。”	0.0	-19.9	0.00	-0.82	0.02	-0.60
S095	“外面的院子真不错。”	0.0	-19.9	0.00	-0.82	0.02	-0.58
S095	“真是一团糟。”	0.0	-19.9	0.00	-0.82	0.02	-0.60
S095	“正在擦干餐具。”	0.0	-19.9	0.00	-0.82	0.02	-0.58
S095	“我觉得她正朝窗 户外看。”	0.0	-19.9	0.00	-0.82	0.02	-0.60
S107	“椅子。”	28.4	+0.4	0.02	-0.83	0.04	-0.69
S080	“看不到其他东西了。”	0.0	-17.3	0.00	-0.90	0.02	-1.00
S080	“没有。”	0.0	-17.3	0.00	-0.90	0.02	-0.94
S080	“没有。”	0.0	-17.3	0.00	-0.90	0.02	-0.98
S086	“有两个杯子和一个茶碟，或者一个盘子吧。”	25.1	-0.1	0.06	-0.91	0.05	-0.14
S107	“xxx 。”	27.9	-0.1	0.02	-0.91	0.18	+0.21
S100	“我不知道 &=笑了。”	32.6	+2.9	0.04	-0.92	0.03	-1.09
S124	“（.）那个小女孩正伸手 去拿一块饼干。”	24.0	+0.9	0.09	-0.93	0.04	-0.55
S125	“嗯哼。”	37.3	+3.7	0.05	-0.96	0.04	-0.18
S081	“女孩在嘲笑她的兄弟 因为他去了.....”	21.5	-0.9	0.10	-0.98	0.42	+1.10
S082	“也许他有 +.....”	28.3	+1.3	0.17	-1.06	0.04	-0.49
S079	“比如说，母亲就在女孩附近 .”	23.3	-1.4	0.16	-1.09	0.45	-0.02
S039	“我该从哪里开始? ”	27.9	-2.1	0.13	-1.13	0.50	-0.57
S086	“食品储藏室上面的罐子里有饼干我想是这样。”	25.4	+0.2	0.03	-1.17	0.02	-0.99
S126	“看起来也像个小孩子。”	27.4	-0.0	0.17	-1.17	0.05	-0.29

（续见下页）

（接上一页）

Spk	话语	基频（半音）		基频方差		响度	
		eGe	spk	eGe	*z*	eGe	*z*
S090	“+< 你.....了吗？”	30.3	+1.1	0.17	-1.20	0.77	+3.73
S135	“她看起来 +//怎么样？”	24.4	+0.0	0.03	-1.23	0.02	-0.86
S126	“看起来里面像是有个小女孩 .”	29.6	+2.1	0.16	-1.25	0.17	+2.83
S124	“男孩正在拿饼干。”	22.4	-0.7	0.06	-1.26	0.04	-0.66
S100	“哦 (.) +...”	17.3	-12.3	0.01	-1.26	0.02	-1.24
S076	“嗯，_很多_事情正在发生 .”	29.9	+1.2	0.23	-1.31	1.43	+0.20
S081	“她正朝窗户外看。”	23.0	+0.7	0.05	-1.31	0.14	-1.36
S125	“还有什么呢？”	30.5	-3.1	0.03	-1.35	0.02	-0.60
S093	“他还在拿饼干。”	30.5	+0.2	0.00	-1.39	0.03	-0.71
S093	“她还在告诉他要 sh@o be 安静下来，这样母亲就赢了.....”	31.2	+0.9	0.00	-1.42	0.03	-0.87
S093	“我想她本来会转身的。”	0.0	-30.3	0.00	-1.42	0.02	-0.96
S126	“我没看到别的东西。”	33.2	+5.8	0.15	-1.43	0.02	-1.22
S139	“嗯。”	33.1	+4.7	0.00	-1.52	0.02	-1.10
S139	“他手里拿着一块饼干。”	0.0	-28.4	0.00	-1.52	0.02	-1.11
S135	“她 +.....”	0.0	-24.4	0.00	-1.59	0.01	-1.02
S089	“嗯。”	35.9	+6.7	0.15	-1.61	0.11	+0.45
S082	“整个地板上到处都是。”	23.0	-4.0	0.08	-2.01	0.08	-0.33
S090	“嗯。”	23.5	-5.7	0.07	-2.02	0.06	-1.03
S082	“男人！”	0.0	-27.0	0.00	-2.85	0.02	-0.59

C 按情绪划分的声学特征

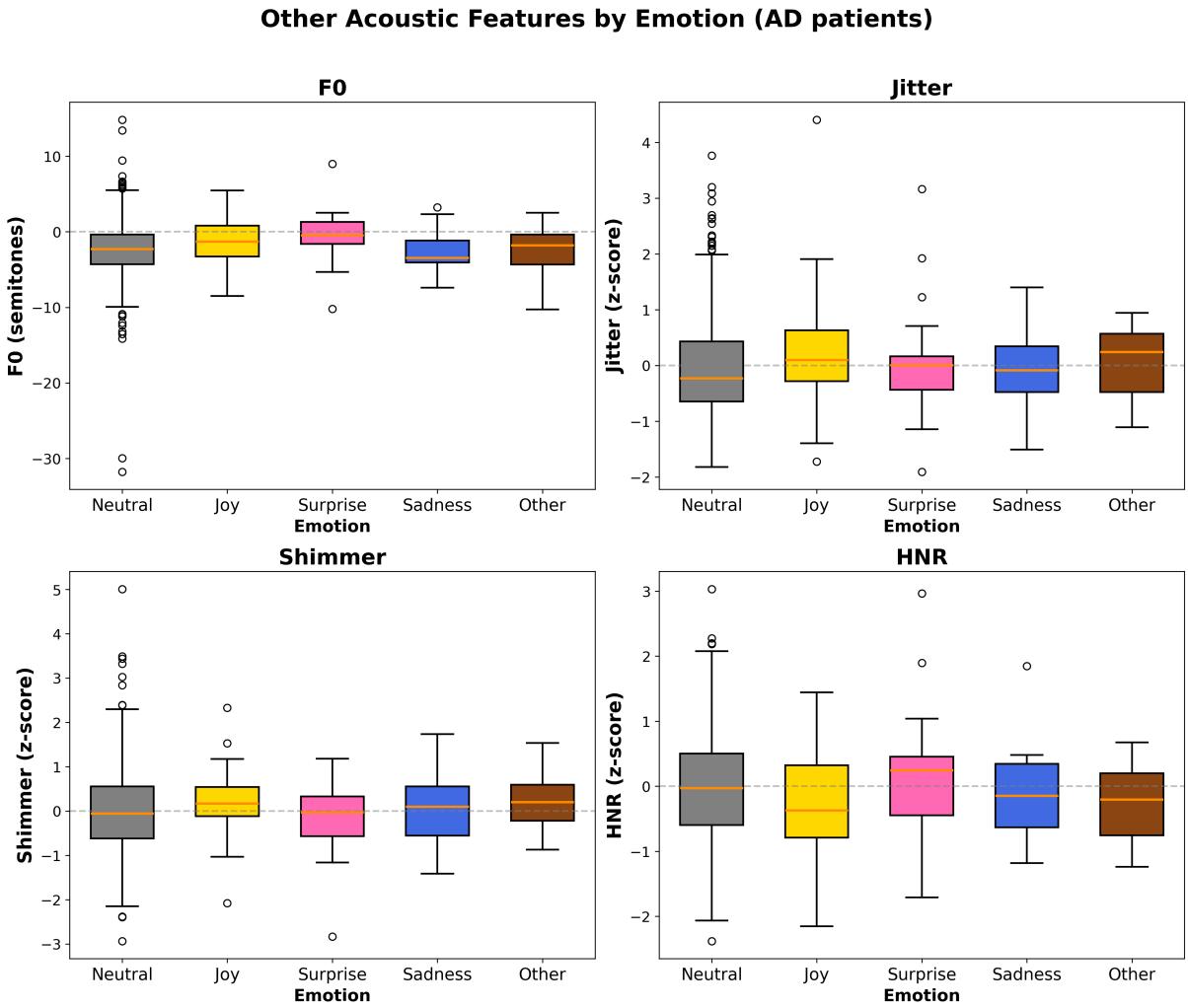


图3：按情绪类别划分的说话人归一化声学特征（仅限阿尔茨海默病患者）。误差条表示95%置信区间。