

T/GZBZ

广州市标准化协会团体标准

T/GZBZ XX—20XX

儿童病理语音结构化采样规范

Specification for structured sampling of pathological speech
in children

（征求意见稿）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

20XX—XX—XX 发布

20XX—XX—XX 实施

广州市标准化协会 发布

目 次

前言..... III

1 范围..... 1

2 规范性引用文件..... 1

3 术语和定义..... 1

4 采样方式..... 1

5 采样过程..... 2

6 采样目标..... 5

附录 A（规范性） 中国构音障碍语音评估系统词表..... 6

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由广州科慧健远医疗科技有限公司提出。

本文件由广州市标准化协会归口。

本文件由广州科慧健远医疗科技有限公司负责具体解释和实施情况收集。

本文件起草单位：广州科慧健远医疗科技有限公司、澳门国际人工智能学会、永讯亚洲有限公司、广州言成智能设备股份有限公司、武汉昱呈星医疗科技有限责任公司、科慧健远（广州）智能设备有限公司、广州鹏远智能设备有限公司。

本文件主要起草人：XXX、XXX、XXX、XXX、XXX、XXX、XXX。

儿童病理语音结构化采样规范

1 范围

本文件规定了儿童病理语音结构化采样的术语和定义、采样方式、采样过程及采样目标等要求。
本文件适用于 2~6 岁儿童病理语音结构化采样。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

下列术语和定义适用于本文件。

3.1

病理语音采样 pathological speech sampling

通过数码音频系统将病变语音声波波形转换成一连串的二进制数据,从而记录下原始模拟声波在某一时刻的状态。

3.2

采样点 sample points

采集2~6岁儿童病理语音的隔音室。

3.3

采样设备 sampling equipment

基于语音采样标准化数据库,对采样的病理语音进行储存的设备。

3.4

采样分析技术 sampling analysis technology

病理语音结构化采样标准涉及一种病理语音采样方法,特别是一种基于 MFCC 特征提取的可以用于人工智能运算的标准化采样方法。

3.5

采样类型辨别 sample type discrimination

在病理语音采样类别中,智能辨别儿童病理语音种类的操作。

4 采样方式

4.1 人工病理语音采样

在采样点,受试者使用普通话或粤语按照附录 A 《中国构音障碍语音评估系统词表》(简称词表)的顺序读取 82 个汉语的语音,检查者通过采样设备进行病理语音采样。

4.2 智能病理语音采样

通过采样分析技术筛选出评估各型儿童病理语音的权重声学特征参数，基于权重声学特征通过构建 ANN 实现 2~6 岁儿童病理语音的采样类型辨别。

5 采样过程

5.1 采样点环境选择

最佳采样点环境是在具有隔音门、吸音岩绵的隔音室内进行，隔音度不小于 40 dB。

5.2 采样设备的基础参数要求

采样设备需要具备 44.1 kHz 的采样率和 16 bit 的音质。

5.3 病理语音采样数据采集

5.3.1 按照词表中的 82 个汉语词汇，对受试者进行录音，采集 82 个汉语音节的病理语音样本。

5.3.2 受试者应端坐，口唇距采样设备约 10 cm，语速自然平稳、音量适中，将词表重复录 2 次，如遇不认识的字检查者应予以告知。

5.4 采样处理

5.4.1 用语音剪辑软件（CoolEdit Pro 2.1）将受试者第一次录音的 82 个目标音档剪切出来。

5.4.2 分类归档，其中单元音 28 个，复合元音 23 个，辅音 21 个，序列语音 10 个。

5.4.3 如果第一次录音出现噪音、干扰、波形超过窗口值、波形提示能量不足等情况，则选用第二次录音数据进行处理。如果再次出现噪音、干扰、波形超过窗口值、波形提示能量不足等情况需重新采样。

5.5 采样分析

5.5.1 概述

通过预加重、分帧、加窗、快速傅里叶变换、三角带通滤波器、滤波代入、扩展分帧等处理过程，提取每个音节的每一个数字语音信号的 MFCC 特征，形成 MFCC 语音库。

5.5.2 预加重

5.5.2.1 预加重的作用：

- a) 使信号的频谱变得平坦，保持在低频到高频的整个频带中，能用同样的信噪比来获得频谱；
- b) 消除发音过程中声带和嘴唇的相互效应，从而补偿语音。

5.5.2.2 将采样点采集的病理语音采样信号通过高通滤波器进行高通预加重处理，其处理公式（1）如下：

$$H(Z) = 1 - \mu z^{-1} \dots\dots\dots (1)$$

式中：

$H(Z)$ ——高通预加重值；

μ ——信噪比， μ 的值为 0.97（可选 0.9~1.0）；

z ——频谱。

5.5.3 分帧

5.5.3.1 规定时间以 25 ms 为一分帧。

5.5.3.2 为了避免相邻两帧的变化过大，两相邻帧之间重叠区域应设置为 10 ms。

5.5.3.3 每一个语音样本取第 13 分帧和第 19 分帧，若语音样本时长不足，补零。

5.5.4 加窗

分帧后将每一帧乘以汉明窗 (Hamming Window)，以增加帧左端和右端的连续性。假设分帧后的信号为 $S(n)$ ，其中 $n = 0, 1, \dots, N-1$ ； $N=13$ 和 $N=19$ ，即帧的大小，那么乘上汉明窗后的信号值为 $x(n)$ ，其公式 (2) 如下：

$$x(n) = S(n) \times W(n) \quad \dots\dots\dots (2)$$

式中：

$x(n)$ ——加窗信号值；

$S(n)$ ——分帧后信号值；

$W(n)$ ——汉明窗。

汉明窗提取公式 (3) 如下：

$$W(n, a) = (1 - a) - a \times \cos \left[\frac{2\pi n}{N-1} \right], 0 \leq n \leq N-1 \quad \dots\dots\dots (3)$$

式中：

$W(n, a)$ ——汉明窗提取值；

n ——分帧后帧的大小；

a ——分帧剔除值， a 值取 0.46；

N ——帧的大小。

5.5.5 快速傅里叶变换

5.5.5.1 对分帧加窗后的各帧信号进行快速傅里叶变换得到各帧的频谱值。

5.5.5.2 对语音信号的频谱值取模平方得到语音信号的功率谱。

5.5.5.3 设定语音信号的离散傅里叶变换 (DFT) 公式 (4) 如下：

$$X_a(k) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi kn/N}, 0 \leq k \leq N \quad \dots\dots\dots (4)$$

式中：

$X_a(k)$ ——离散傅里叶变换；

N ——傅里叶变换点数；

n ——分帧；

k ——离散傅里叶变换点数。

上式中 $x(n)$ 为输入的语音信号， N 表示傅里叶变换的点数。

5.5.6 三角带通滤波器

5.5.6.1 将 $X_a(k)$ 通过一组 24 个三角形滤波器，该滤波器中心频率指定为 $f(m)$ ， $m=1, 2, \dots, 24$ 。各 $f(m)$ 之间的间隔随着 m 值的减小而缩小，随着 m 值的增大而增宽，如图 1 示。

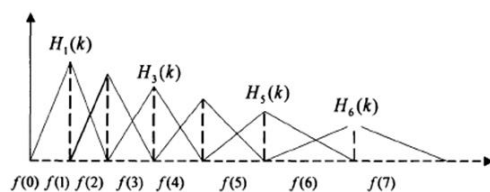


图1 Mel 频率滤波器组模式图

5.5.6.2 三角带通滤波器对病理语音采样获取的样本频谱数据进行处理的作用为：

- 对病理语音采样获取的样本频谱进行平滑化处理，并对频谱起到消除谐波的作用；
- 降低后续采样分析中对样本频谱数据的采样样板辨别的算法计算量。

5.5.6.3 计算数频率值的三角带通滤波器公式（5）如下：

$$H_m(k) = \begin{cases} 0 & , k < f(m-1) \\ \frac{2(k-f(m-1))}{(f(m+1)-f(m-1))(f(m)-f(m-1))}, & f(m-1) \leq k \leq f(m) \\ \frac{2(f(m+1)-k)}{(f(m+1)-f(m-1))(f(m)-f(m-1))}, & f(m) \leq k \leq f(m+1) \\ 0 & , k \geq f(m+1) \end{cases} \quad \dots\dots\dots (5)$$

式中：

- $H_m(k)$ ——对数频率值；
 k ——三角滤波器组值；
 $f(m)$ ——三角滤波器中心频率；
 m ——三角滤波器个数。

上式中， $\sum_{m=0}^{M-1} H_m(k) = 1$ 。

5.5.7 滤波代入

每个滤波器组输出的对数频率值， $H_m(k)$ 代入公式（6）：

$$s(m) = \ln(\sum_{k=0}^{N-1} |X_a(k)|^2 H_m(k)), 0 \leq m \leq M \quad \dots\dots\dots (6)$$

离散余弦变换（DCT）， $S(m)$ 代入公式（7）：

$$C(n) = \sum_{m=0}^{N-1} s(m) \cos\left(\frac{\pi n(m-0.5)}{M}\right), n = 1, 2, \dots, L \quad \dots\dots\dots (7)$$

式中：

- $S(m)$ ——每个滤波器组输出的对数频率值；
 N ——傅里叶变换点数；
 k ——离散傅里叶变换点数；
 $X_a(k)$ ——离散傅里叶变换；
 $H_m(k)$ ——对数频率值；
 L ——MFCC 系数阶数；
 $C(n)$ ——每一分帧 MFCC 值。

上式 L 指 MFCC 系数阶数，通常取 12-16，指定 13 和 19； M 是三角滤波器个数，指定为 24； $C(n)$ 为每一分帧的 MFCC 值。13 阶和 19 阶各分帧连接获得 2 组 MFCC 值入库（A 组、B 组）。

5.5.8 扩展分帧

将三角滤波器中获得的共振峰 F_1 、 F_2 和 F_0 各中点值，作为一个分帧加入 A 组、B 组，获得 2 组 MFCC 入库，即 C 组、D 组

5.5.9 MFCC 语音库

5.5.9.1 MFCC 语音库包括病理语音采样标准化数据和采样类型辨别结构化数据。

5.5.9.2 病理语音采样标准化数据：82 个音节样本的每一个音节数据经预处理后分别是 13 帧、19 帧、13+3 帧、19+3 帧的标准化数据入 A、B、C、D 四组。

5.5.9.3 采样类型辨别结构化数据：将标准化数据分别录入元音与声调、辅音、复合元音、序列数字等四个子库的 A、B、C、D 组。

6 采样目标

6.1 明确儿童病理语音与正常儿童及组间的声学特征差异与规律

6.1.1 采样点通过对采样设备进行病理语音采样采集到的样本进行处理和分析后，将样本的每个音节处理为标准化和结构化数据，以适应病理语音库构建、大数据语音分析和人工智能运算的需求。

6.1.2 通过采样处理后提取到的共振峰、基频、时长、塞音 (VOT) 等相关的声学特征参数，并分析计算出病理语音样本的标准化和结构化数据中具备元音三角 (VSA)、元音发音指数 (VAI)、共振峰集中率 (FCR)、元音椭圆离散度等声学参数，应不少于 25 种。

6.1.3 通过大数据语音分析和人工智能运算，可完整分析出 2-6 岁儿童病理语音与正常语音的声学特征差异和变化规律。

6.2 评估各型儿童病理语音的权重声学特征参数

6.2.1 根据儿童病理语音和正常儿童，分别建立 5 个结构化语音数据集。每个数据集中每个样本语音包括 25 个子集（以下用 $V_1, V_2, \dots, V_{25}$ 表示）。

6.2.2 采样类型辨别能通过数据挖掘的方法选择与优化用于评估 2-6 岁儿童病理语音特征声学参数。

6.2.3 根据病理语音采样预估数据特征选取 Lasso 分析，将每种类型的儿童病理语音进行二分类变量 ($C_1, C_2, \dots, C_4$) 分类，为采样类型辨别提供数据基础。

6.2.4 通过 LASSO 分析，明确影响采样类型辨别中的 25 种变量因素的比例权重，并精确确定用于儿童病理语音评估的综合权重因变量，即权重声学特征，提高信效度。

6.3 基于权重声学特征通过构建 ANN 实现儿童病理语音类型辨别

6.3.1 将获取到的儿童病理语音样本的权重因变量（例如， $V_1, V_8, \dots, V_{22}$ ）为输入变量，ANN 四分类疾病作为输出变量，输入 ANN 神经网络，获取儿童病理语音权重变量。

6.3.2 基于声学特征的 ANN 模型对儿童病理语音识别的效率与影响机制，提高权重声学特征参数进行人工智能、客观评估儿童病理语音疾病的信效度。

附 录 A
(规范性)
中国构音障碍语音评估系统词表

A.1 中国构音障碍语音评估系统词表包括元音、声调、系列语言、元音变化、部分韵母和辅音等内容，具体内容见表 A.1、表 A.2 和表 A.3 的规定。

表 A.1 规定了文件中使用的中国构音障碍语音评估系统词表的元音、声调、系列语言。

表 A.1 中国构音障碍语音评估系统词表（元音、声调、序列语言）

序号	字	声母	韵母	声调
1	八	b	a	1
2	拔	b	a	2
3	把	b	a	3
4	爸	b	a	4
5	逼	b	i	1
6	鼻	b	i	2
7	笔	b	i	3
8	必	b	i	4
9	都	d	u	1
10	读	d	u	2
11	赌	d	u	3
12	杜	d	u	4
13	哥	g	e	1
14	隔	g	e	2
15	葛	g	e	3
16	个	g	e	4
17	波	b	o	1
18	脖	b	o	2
19	跛	b	o	3
20	簸	b	o	4
21	淤	y	ü	1
22	鱼	y	ü	2
23	雨	y	ü	3
24	玉	y	ü	4
25			a	1
26			a	2
27			a	3
28			a	4
29	1	y	i	1
30	2		er	4
31	3	s	an	1

表 A.1 中国构音障碍语音评估系统词表（元音、声调、序列语言）（续）

序号	字	声母	韵母	声调
32	4	s	i	4
33	5	w	u	3
34	6	l	iu	4
35	7	q	i	1
36	8	b	a	1
37	9	j	iu	3
38	10	sh	i	2

表 A.2 规定了文件中使用的中国构音障碍语音评估系统词表的元音变化和部分韵母。

表 A.2 中国构音障碍语音评估系统词表（元音变化、部分韵母）

序号	字	声母	韵母	声调
1	掰	b	ai	1
2	虾	x	ia	1
3	包	b	ao	1
4	瓜	g	ua	1
5	丢	d	iou	1
6	龟	g	uei	1
7	杯	b	ei	1
8	憋	b	ie	1
9	标	b	iao	1
10	边	b	ian	1
11	班	b	an	1
12	宾	b	in	1
13	奔	b	en	1
14	帮	b	ang	1
15	冰	b	ing	1
16	崩	b	eng	1
17	锅	g	uo	1
18	光	g	uang	1
19	关	g	uan	1
20	沟	g	ou	1
21	乖	g	uai	1
22	靴	x	üe	1
23	兄	x	iong	1

表 A.3 规定了文件中使用的中国构音障碍语音评估系统词表的辅音。

表 A.3 中国构音障碍语音评估系统词表（辅音）

序号	辅音类型	字	声母	韵母	声调
1	不送气塞音	八	b	a	1
2	送气塞音	趴	p	a	1
3	不送气塞音	搭	d	a	1
4	送气塞音	他	t	a	1
5	不送气塞音	嘎	g	a	1
6	送气塞音	喀	k	a	1
7	不送气塞擦音	机	j	i	1
8	送气塞擦音	七	q	i	1
9	不送气塞擦音	知	zh	i	1
10	送气塞擦音	吃	ch	i	1
11	不送气塞擦音	资	z	i	1
12	送气塞擦音	疵	c	i	1
13	清擦音	发	f	a	1
14	清擦音	哈	h	a	1
15	清擦音	西	x	i	1
16	清擦音	师	sh	i	1
17	清擦音	思	s	i	1
18	浊擦音	日	r	i	4
19	鼻音	妈	m	a	1
20	鼻音	那	n	a	1
21	边音	拉	l	a	1

A.2 中国构音障碍语音评估系统词表说明：

- 表 A.1 由 38 个汉字构成，覆盖 a、i、u、o、e、ü 等 6 个元音的四种声调，包含 /a/ 不在汉字中单独发音的四种声调；
- 为了适应少数左脑损伤不能表达汉字，但是序列语言保留能读 1-10 的数字的病人，表 A.1 设计了 1-10 数字发音；
- 表 A.2 由 23 个汉字构成，包括 13 个复韵母的第一声调发音和 10 个鼻韵母的第一声调发音；
- 根据运动性构音障碍的相关肌肉存在协调障碍的问题，为了研究元音变化的异常，表 A.2 包含了 a-i-u 这三个在元音图上位置相距最远的元音间的变化，a-i (/ai/)、i-a (/ia/)、i-u (/ (y) iu/)、u-i (/wei/)、a-u (/ao/)、u-a (/wa/);
- 表 A.3 由 21 个汉字构成，包括 21 个声母，并按塞音、塞擦音、擦音、鼻音、边音顺序排列。