**系统评价算法测试**

系统评价算法测试主要有三个部分，分别是音符级评价测试、小节评价测试和整曲评价测试。

* 音符级评价测试

多基音检测系统通用的评价标准之一为note level F-measure，计算方法如下

其中，nCorr表示检测正确的音符总数，nRef表示正确结果中的音符总数，nTot表示检测到的音符总数。表1为MAPS数据集的音符级评价结果,采样率为44100，。

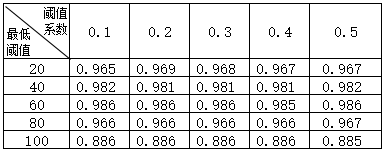


表1 甘老师数据集的音符级评价结果表（fs=44100）

本数据集的特点为音量大，同时弹奏的音符少，节奏慢。从测试结果表中可以看到，在最低阈值不超过80的情况下结果均超过了95%，最高达到了98.6%；阈值系数对评价结果的影响不大，在1%以内。

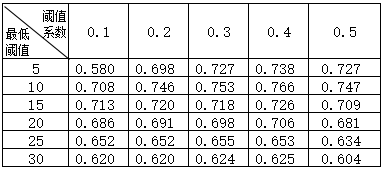


表2 MAPS数据集的音符级评价结果表（fs=44100）

MAPS数据集的特点为音量小，同时弹奏的音符多，节奏快。从结果中可以看到，在最低阈值不超过20、阈值系数设定在0.2～0.5时结果较好。