

Шаг 1 - какой метод отвечает на ваш вопрос?

MOS	«Насколько хорош клип, 1-5?» Абсолютная оценка речи (ITU-T P.800; P.808 для краудсорсинга). Каждый клип оценивают отдельно.
MUSHRA	«Ранжируй кодеки, 0-100.» Промежуточное качество (ITU-R BS.1534-3). Все системы сразу + скрытый эталон + якорь-ФНЧ.
A/B	«Что предпочтительнее?» Доля предпочтений (%). Дёшево и понятно - но сначала выровняйте громкость, иначе мерите громкость.
ABX	«Слышно ли различие вообще?» X тайно равен A или B; угадайте. Значимо, только если биномиальный тест бьёт 50% угадывания.

Шаг 2 - сколько слушателей, какая статистика?

Метод	Слушателей (валидных)	Стат. тест	Результат
MOS (P.800)	24+ лаб / сотни крауд	среднее + ДИ; ANOVA	балл 1-5
MUSHRA (BS.1534-3)	~15+ после отсева	среднее + ДИ; RM-ANOVA	балл 0-100
A/B предпочтение	20-40+	биномиальный / знаковый	доля %
ABX	1 обуч. + много проб	биномиальный vs 50%	p-значение

Факты, на которых спотыкаются

- MUSHRA = Multiple Stimuli with Hidden Reference and Anchor. Скрытый эталон задаёт верх (около 100); якорь-ФНЧ 3.5 кГц задаёт низ.
- Отсев в MUSHRA убирает слушателя, оценившего скрытый эталон ниже 90 более чем в 15% заданий - набирайте больше 15, чтобы после отсева осталось 15.
- MOS - не переносимая линейка: 4.2 имеет смысл только внутри одного теста. Не цитируйте «голый» MOS без контекста теста.
- ABX: большинство - не доказательство. 12 из 16 значимо ($p = 0.038$); 10 из 16 - нет ($p = 0.23$). Точность растёт как $\sqrt{\text{слушателей}}$.
- BS.1116-3 (ABC/HR) - «брат» MUSHRA для почти неотличимых, аудиофильских различий; MUSHRA - для слышимого, но приемлемого диапазона.

Защитимый тест - дисциплина

- ☐ Сначала выровняйте все клипы по одному LUFs - громче надёжно «побеждает» лучше.
- ☐ Зафиксируйте метод и число слушателей до сбора первой оценки.
- ☐ Рандомизируйте и балансируйте порядок клипов, чтобы убрать смещение порядка.
- ☐ Отсеивайте слушателей (скрытый эталон / проверки внимания); невнимательных - вон.
- ☐ Сообщайте доверительные интервалы и тест значимости - «голое» среднее это догадка.