

О высоте гласных в речевом сигнале

А.И.Егоров, В.В.Дубровский

Институт динамики систем и теории управления СО РАН

egorov@ccsoan.irkutsk.su

Как отмечалось в работе [1], анализ речи по ее описаниям, при выборе которых практически полностью игнорируется определяющее участие головного мозга в порождении и восприятии речевого сигнала (РС), недостаточно обоснован. В этом отношении более перспективен бионический подход, базирующийся на некоторых общих принципах обработки сенсорной информации, предположительно используемой человеком. С позиций бионического подхода одной из характеристик РС, полезной для экономного и адекватного описания слуховых образов речи, является высота. Разными авторами для интерпретации тех или иных экспериментальных фактов (например, восприятия высоты резидуальных звуков) предлагались различные варианты временных и спектральных методов оценки высоты. Да и сама высота называлась по-разному: фактической, спектральной, резидуальной, виртуальной и так далее. В приложениях же высоту гласных, как правило, продолжают приближенно оценивать по периоду основного тона РС. Этот метод относительно прост в реализации, но не учитывает многие экспериментальные факты и, следовательно, не гарантирует от ошибок в оценке высоты. Так, например, при исследовании слухового восприятия высоты гласных часто приходится иметь дело с ситуациями, когда период основного тона РС не коррелирует с ощущением высоты. В сообщении предложен вариант спектрально-временного подхода к оценке высоты, который позволяет достичь лучшего согласования значений, полученных расчетным и экспериментальным путем. Специфика нашего подхода состоит в попытке учесть влияние на ощущение высоты процессов взаимодействия соседних тональных компонент в речевых созвучиях.

Известно, что при восприятии созвучий, тональные составляющие которых не связаны гармоническим соотношением, слуховая система человека может выделить несколько простых сигналов. Причем ощущение высоты каждого из таких сигналов определяется их частотой. Для гласных, спектр которых может быть представлен в виде гармонического комплекса с частотными компонентами $F_i = i \cdot F_0$, (где i - целое, F_0 - частота основного тона) характерно слуховое восприятие созвучия в виде единого слухового образа, обладающего некоторой высотой.

Рассмотрим структуру высоты слуховых образов сигналов, представленных гармоническими комплексами. Пусть стимул, воздействующий на слуховую систему, представлен двумя

тонами с соответствующими частотами F_1 и F_2 , а интервал частот $(F_2 - F_1)$ меньше ширины частотной группы на участке спектра, к которому относятся F_1 и F_2 . При этом, как следует из [2], в слуховой системе образуется область двухтонового возбуждения, воспринимаемая в виде единого слухового образа. В этой области базилярная мембрана внутреннего уха одновременно участвует в двух колебательных процессах: высокочастотном и низкочастотном (биения с разностной частотой

$(F_2 - F_1) = F_0$. Интенсивность биений зависит от интенсивности тонов в стимуле. Отметим также, что при перемещении частот стимула в более низкочастотную область спектра, в которой условие частотного разрешения тонов выполняется лучше, чем в высокочастотной, область двухтонового возбуждения, по-видимому, не исчезает. Но интенсивность биений (при сохранении интенсивности исходных тонов) уменьшается. Приведенная информация позволяет сделать предположение о том, что ощущение высоты, возникающее при восприятии двухтонового сигнала, формируется из двух ощущений: "высокой" и "низкой" высоты. "Высокая" составляющая высоты связана с высокочастотным колебательным процессом на мембране, а "низкая" - с биениями с разностной частотой. Можно также предположить, что итоговая количественная оценка высоты единого слухового образа, соответствующего восприятию всего гармонического комплекса, зависит от соотношения громкостей слуховых ощущений "высокой" и "низкой" высоты.

С целью проверки выдвинутых предположений авторами выполнен большой цикл экспериментальных исследований. Стимулы, использованные в экспериментах, формировались на компьютере программным путем. Спектры стимулов представляли гармонические комплексы. Значения частот и интенсивностей спектральных составляющих стимулов задавались и изменялись в соответствии с задачами, решаемыми в рамках конкретных экспериментов. В каждом эксперименте на громкоговоритель поочередно выдавалось два стимула. И затем аудиторы пытались дать сравнительные оценки высот воспринимаемых созвучий. В качестве примера можно привести следующий. В очередном эксперименте использовалось три созвучия, каждое из которых состояло из пяти тональных составляющих:

созвучие 1 (500гц, 60дб; 600гц, 60дб; 800гц, 75дб; 1000гц, 60дб; 1100гц, 60дб);

созвучие 2 (500гц, 65дб; 600гц, 65дб; 800гц, 75дб; 1000гц, 65дб; 1100гц, 65дб);

созвучие 3 (500гц, 55дб; 600гц, 55дб; 800гц, 75дб; 1000гц, 55дб; 1100гц, 55дб).

От аудиторов были получены следующие сравнительные оценки высоты созвучий: $H_1 > H_2$; $H_1 < H_3$; $H_2 < H_3$. Здесь H_1, H_2, H_3 - высота, соответственно, 1, 2 и 3 созвучия. Таким образом, $H_2 < H_1 < H_3$.

Анализ результатов всех проведенных экспериментов позволил сделать вывод о том, что они в целом подтвердили правомерность наших предположений и позволили получить следующую эмпирическую формулу для оценки высоты гласных: $H = F_0 + 0,1 * (Z_m - F_0) * (G_m / G_{bi})$, где H - высота созвучия; Z_m - высота (по шкале мел) тональной составляющей, обеспечивающей максимальный вклад в громкость созвучия; G_m - громкость (в сонах) тональной составляющей с высотой Z_m ; G_{bi} - общая громкость (в сонах) разностных тонов в созвучии. Из формулы видно, что при $G_m = 0$, $H = F_0$. А при $(G_m / G_{bi}) = 10$, $H = Z_m$. Следовательно, в зависимости от соотношения интенсивностей спектральных составляющих, используемых при расчете G_m и G_{bi} , значение H может изменяться от F_0 до Z_m .

При расчете G_m и G_{bi} частично использовались результаты, изложенные в [3]. Вклад в громкость, связанный с нелинейными квадратическими искажениями и комбинационными тонами третьего порядка с частотой $2F_1 - F_2$, в расчетах G_{bi} не учитывался.

В заключение заметим, что сравнительный анализ экспертных оценок высоты гласных из РС нескольких дикторов с результатами расчетов по предложенной в сообщении формуле, свидетельствовал об их удовлетворительном согласии.

Список литературы

Egorov A.I., Dubrowsky V.V. On Analysis of Hearing Image Speech

Signals, Web Journal of Formal, Computational & Cognitive Linguistics, <http://www.ksu.ru/kazan/science/fccl/index.html>.

Позин Н.В., Любинский И.А. и др. Элементы теории биологических анализаторов.- М.: Наука, 1978.

Цвиккер Э., Фельдкеллер Р. Ухо как приемник информации.-М.: Связь, 1971.