



Акустические признаки гласных звуков с негармонической структурой

Уплисова К.О.

Санкт-Петербургский государственный университет.
Россия, 199034, Санкт-Петербург, Университетская наб., 7/9.
Тел. (812) 328-97-06. E-mail: sehmet@fromru.com

Человек различает звуки речи независимо от того, произнесены ли они человеком или птицей, голосом или шёпотом, пропеты или синтезированы. Однако в зависимости от способа генерации, акустические характеристики фонем могут существенно отличаться. Шёпотная речь и звуки, имитируемые птицами, по большей части, не являются гармоническими, а обладают широкополосной структурой. Исследование мощности спектральных компонентов в различных частотных полосах для гласных звуков серого попугая и шёпотной речи выявило их относительное сходство. Для определения ключевых частот, необходимых для идентификации звуков «у», «о», «а», «и», исследована возможность идентификации синтезированных одноформантных шумоподобных звуков: фрагментов белого шума с усиленной на 35 дБ узкой полосой частот. Если полоса находилась в диапазоне от 200 до 500 Гц, звук идентифицировался как «у», 700-900 Гц — «о», 1000–1700 — «а», и если усиленная по амплитуде частота превышала 1900 Гц, звук идентифицировался как «и». Таким образом, на основании первого и единственного максимума в спектре возможно идентифицировать четыре категории гласных. При сравнении значений выраженных спектральных компонентов гласных звуков серого попугая и шёпотных гласных в диапазонах частот, соответствующих полученным при идентификации синтезированных звуков, оказалось, что у всех звуков присутствуют выраженные спектральные компоненты в соответствующих зонах. В связи с этим можно предположить, что для идентификации гласных категорий «у», «о», «а», «и» необходимо и достаточно наличия выраженных спектральных компонентов в определённой частотной зоне. Выраженные спектральные максимумы вне этих зон могут не оказывать влияния на идентификацию категории гласного звука.

ВВЕДЕНИЕ

Человек различает звуки речи независимо от того, произнесены ли они человеком или птицей, голосом или шёпотом, пропеты или синтезированы. Однако в зависимости от способа генерации акустические характеристики фонем могут существенно отличаться. Речь, генерируемая голосом взрослого человека, имеет частоту основного тона от 100 до 250 Гц и обладает выраженной гармонической структурой. Наиболее энергетически выраженные точки огибающей спектра считаются формантами, и на основании частот первых двух формант возможно разделить все категории гласных звуков [1, 2, 3]. Получены данные [4, 5, 6, 7], свидетельствующие о том что, что при значениях частот основного тона, не превышающих 240 Гц, у большинства гласных выделяется типичная формантная структура, и на двухформантной плоскости наблюдается характерное разделение областей

их представительства. По мере возрастания частоты основного тона, как формантная характеристика звуков, так и их оценка по спектральным максимумам уже не позволяют разделить гласные, относящиеся к разным фонетическим категориям, поскольку энергетически выраженными оказываются три-четыре гармонических составляющих, включая и первую, соответствующую частоте основного тона. В то же время анализ относительной амплитуды первых четырёх гармонических составляющих показал, что эти характеристики меняются в зависимости от значений частоты основного тона по-разному у разных гласных. Таким образом, оценка по признаку относительной амплитуды спектральных компонентов позволяет не только выявить различия между гласными [а], [о], [у] при высоких значениях частоты основного тона, но и разделить гласные речевого диапазона при неопределённости их формантной картины независимо от формы речи (в том числе и гласные [а] и [и], имитируемые скворцом-майной) [8]. Более того, звуки с типичной формантной структурой также могут быть описаны с использованием данного признака.

С другой стороны, шёпотная речь и речь, имитируемая говорящими птицами, вообще лишена основного тона и не имеет гармонической структуры. Выраженные по амплитуде спектральные компоненты, которые можно принять за форманты, не всегда находятся в диапазонах, соответствующих значениям первых двух формант обычной голосовой речи [9]. Таким образом, возможность разделения гласных звуков на основании значений первых двух формант является лишь частным случаем. Возникает вопрос, существует ли единый, универсальный признак, на основании которого возможно разделение фонетических категорий звуков независимо от способа их генерации, или же для каждого типа звукогенерации существует своя собственная система различительных акустических признаков.

Для негармонических гласных говорящих птиц учёт амплитудных характеристик спектральных компонентов не даёт результата. Ключевыми признаками являются частотные значения формант, однако если в звуке присутствует несколько формантных областей, вклад их в идентификацию категории фонемы не одинаков. Было выявлено [10], что если в спектре звука присутствуют выраженные спектральные компоненты в диапазоне 1300–1700 Гц, этот звук будет с высокой вероятностью идентифицирован как «а», независимо от того, присутствуют ли выраженные спектральные компоненты на низких и высоких частотах. Если частоты основных энергетически выраженных компонентов находятся ниже 1000 Гц, то звук будет идентифицирован как «о», независимо от присутствия выраженных спектральных компонентов выше 1700 Гц, но не в случае их наличия в диапазоне 1300–1700 Гц. В случае наличия энергетически выраженных спектральных компонентов с частотой ниже 500 Гц и при отсутствии выраженных спектральных компонентов на более высоких частотах звук будет идентифицироваться как «у». Если выраженные спектральные компоненты будут присутствовать начиная с 1700 Гц, звук будет идентифицироваться как «и». В связи с тем, что способ звукогенерации у птиц существенно отличается от человеческого, возникает вопрос, какими характеристиками будут обладать гласные звуки человеческой речи с негармонической структурой — шёпотные гласные, и возможно ли во всех негармонических гласных выявить единый различительный признак, характерный для той или иной фонетической категории звука.

МЕТОДИКА

Для исследования акустических характеристик шёпотных гласных были получены аудиозаписи 160 юношей и девушек (18–20 лет). Перед студентами была поставлена задача шёпотом произнести звуки категорий «а», «о», «у», «и», «ы», «э» по четыре раза каждый, с каждым разом пытаясь увеличить высоту голоса. Из полученных записей было выделено 756 звуков, произнесённых без колебания голосовых складок. Выделение звуков осуществлялось на слух, по однородности динамической узкополосной спектрог-

раммы (1024 фильтра). Нормализованные звуки были прослушаны тремя опытными аудиторами, и в дальнейшем анализировались те звуки, которые получили единогласную оценку. 152 звука были отнесены к категории «а», 68 — «о», 60 — «у», 147 — «и».

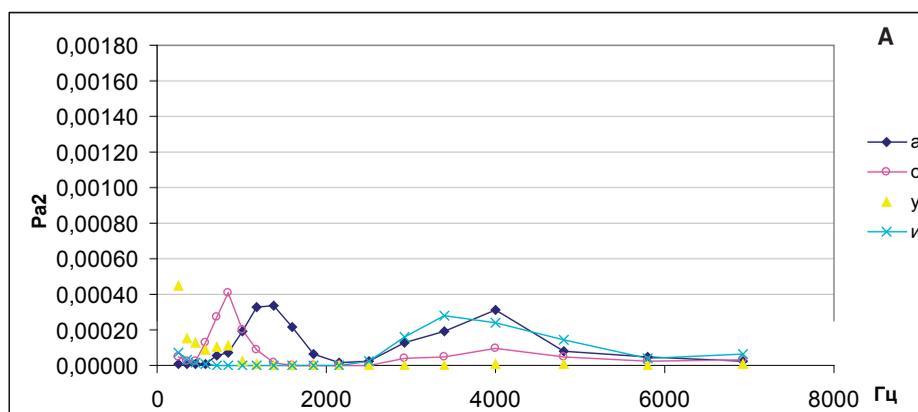
Из аудиозаписи имитации речи серым попугаем Кузей были таким же образом выделены гласные и подвергнуты перцептивному анализу. Звуки прослушивались 25–30 студентами и относились ими к одной из категорий гласных русского языка. Достоверно было идентифицировано 69 «а», 8 «о», 27 «у», 78 «и».

Анализ звуков осуществлялся с помощью программы Praat. С помощью функции фильтрации (pass Hann band) у каждого звука последовательно сохранялись спектральные компоненты в частотном диапазоне, соответствующим значениям частот критических полос слуха человека [11] и определялась их мощность (Pa^2).

Для перцептивного эксперимента была проведена следующая модификация: у серии 500 мс фрагментов белого шума последовательно усиливалась определённая узкая частотная полоса (центральная частота усиливалась на 35 дБ, ширина подъёма и спада области составляла 100 Гц в одной серии и 200 Гц в другой — треугольный фильтр) в диапазоне от 200 Гц до 4 кГц с шагом 100 Гц. Звуки идентифицировались 25 аудиторами. Сначала им предъявлялись четыре модифицированных звука, ранее чётко идентифицировавшихся как «а», «о», «у», «и». После того как аудиторы определяли эти звуки как различные фонемы, перед ними ставилась задача прослушать серию звуков и обозначить, звуки каких категорий они слышат. Достоверность идентификации определялась по биномиальному критерию.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Для 152 звуков «а», 68 — «о», 60 — «у», 147 — «и», произнесённых шёпотом, были получены значения мощности спектральных компонентов в частотных полосах, соответствующих критическим полосам слуха человека (100–200, 200–300, 300–400, 400–510, 510–630, 630–770, 770–915, 915–1080, 1080–1260, 1260–1480, 1480–1720, 1720–1990, 1990–2310, 2690–3125, 3125–3675, 3675–4350, 4350–5250, 5250–6350, 6350–7500 Гц), и определена медиана значений в каждой критической полосе. Анализ выявил, что максимальные значения мощности спектральных компонентов (рис. 1А) для звука «у» находятся в диапазоне 200–300 Гц, «о» — 770–915 Гц, «а» — 1080–1480 Гц и 3675–4350 Гц, «и» — 3125–4350 Гц.



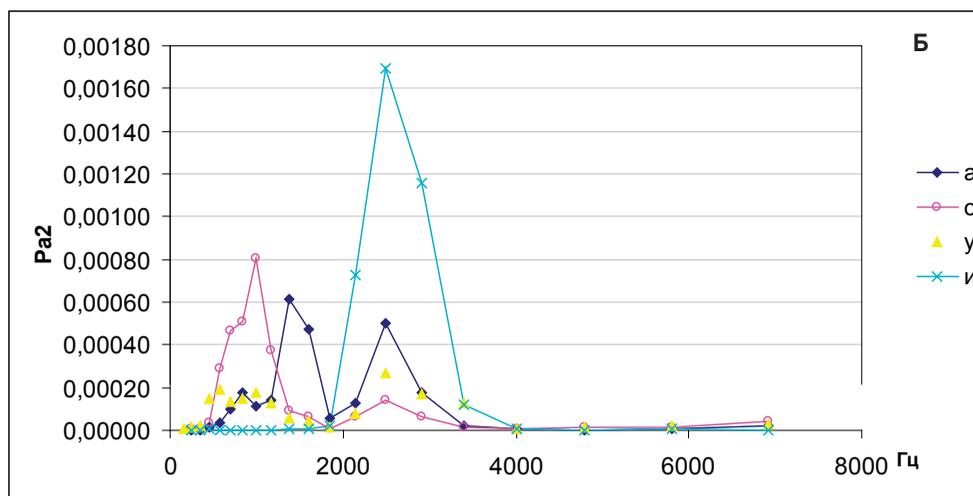


Рис. 1. Мощность спектральных компонентов для гласных различных категорий, произнесённых шёпотом (А) и говорящим попугаем (Б).

По горизонтальной оси — частота, Гц, по вертикальной — мощность, Pa²

Звуки «у» серого попугая характеризовались значительной нестабильностью значений спектральных компонентов (рис. 1Б), которые варьировали в диапазоне 510–1080 Гц, звуки «о» характеризовались максимальными значениями мощности в области 915–1080 Гц, «а» — 1260–1480 и 2310–2690 Гц, «и» — 2310–2690 Гц. В целом использование мощности спектральных компонентов подтвердило данные полученные для амплитудных значений спектральных компонентов гласных звуков говорящих птиц [9]. У звука «а» по сравнению с «о», присутствовали выраженные спектральные компоненты в диапазоне 1260–1480 Гц, у звука «о» по сравнению со звуком «у» присутствовали выраженные компоненты в диапазоне 770–1080 Гц, а звук «и» характеризовался максимумами на частотах 1990–3125 Гц.

Значения первого максимума гласного «а» шёпотной и имитационной речи совпали и оказались на частотах 1260–1380 Гц. Это соответствует данными модификации звука «а» попугаев, свидетельствующим о том, что удаление спектральных компонентов на этих частотах существенно влияют на идентификацию категории «а» [10]. Звуки «о», имитируемые серым попугаем, отличались значениями максимумов, чуть превышающими значения шёпотного «о» (915–1080 Гц против 770–915). Наибольшие различия наблюдались для значений максимумов звука «у», которые составили 510 Гц для имитационного звука и 200–300 Гц для шёпотного. Однако стоит отметить, что звук «у» и у птиц и в шёпотной речи имеет нестабильные характеристики огибающей спектра и идентифицируется значительно хуже, чем все остальные звуки. Что касается звука «и», то максимум имитационного звука располагается существенно ниже, чем шёпотного (2310–2690 против 3125–3675 Гц). При этом значения первых максимумов шёпотных гласных «о», «а», «и» сходны со значениями вторых формант голосовых звуков.

С целью выяснения вопроса о необходимости первой форманты для идентификации гласных звуков, а также уточнения ключевых частотных диапазонов для звуков различных категорий была проведена следующая модификация: в 500 мс фрагментах синтезированного белого шума по амплитуде усиливалась узкая частотная полоса и от звука к звуку смещалась шагами по 100 Гц от 200 до 4000 Гц. Подобные звуки идентифицировались как «у» при нахождении максимума в области от 200 до 500–600 Гц (в зависимости от ширины полосы), от 700 до 800–900 Гц как «о», от 1000–1100 Гц до 1600–2000 Гц как «а», все звуковые модификации с максимумами, превышающими 1800–2100 Гц, однозначно оценивались как фонема «и».

Таким образом для идентификации звуков «у», «о», «а», «и» необходимо и достаточно одного спектрального максимума, находящегося в определённом частотном диапазоне. Эти данные совпадают с результатами работы по оценке модифицированных гласных звуков попугаев [10], в которой было выявлено, что если в спектре звука присутствуют выраженные спектральные компоненты в диапазоне 1300–1700 Гц, этот звук будет с высокой вероятностью идентифицирован как «а», независимо от того, присутствуют ли выраженные спектральные компоненты на низких и высоких частотах. Если частоты основных энергетически выраженных компонентов находятся ниже 1000 Гц, то звук будет идентифицирован как «о», независимо от присутствия выраженных спектральных компонентов выше 1700 Гц, но не в случае их наличия в диапазоне 1300–1700 Гц. В случае наличия энергетически выраженных спектральных компонентов с частотой ниже 500 Гц и при отсутствии выраженных спектральных компонентов на более высоких частотах звук будет идентифицироваться как «у». Если выраженные спектральные компоненты будут присутствовать начиная с 1700 Гц, звук будет идентифицироваться как «и».

ВЫВОДЫ. Акустическая структура шёпотных гласных в значительной степени сходна со структурой гласных звуков, имитируемых птицами. Максимальные значения мощности спектральных компонентов шёпотного звука «у» находятся в диапазоне 200–300 Гц, «о» — 770–915 Гц, «а» — 1080–1480 Гц и 3675–4350 Гц, «и» — 3125–4350 Гц. Синтезированные одноформантные шумоподобные звуки идентифицируются как «у» при нахождении максимума в области от 200 до 600 Гц, от 700 до 900 Гц как «о», от 1100 Гц до 1600 Гц как «а», все звуковые модификации с максимумами, превышающими 2100 Гц, однозначно оцениваются как «и».

ЛИТЕРАТУРА

1. Фант Г. Анализ и синтез речи. Новосибирск: Наука, 1970. 166 с.
2. Фланаган Дж. Анализ, синтез и восприятие речи. М.: Связь, 1968. 396 с.
3. Peterson G.E., Barney H.L. Control methods used in a study of the vowels // Journal of the Acoustical Society of America. 1952. V. 24. P. 175–184.
4. Куликов Г.А. и др. Характеристики гласноподобных звуков детей первого полугодия жизни // Доклады Академии Наук. 1999. Т. 368, № 6. С. 843–845.
5. Андреева Н.Г., Куликов Г.А. Характеристики вокальных гласных при разной частоте основного тона // Рос. Физиол. Журн. им. И.М. Сеченова. 2003. Т. 89, № 6. С. 715–724.
6. Андреева Н.Г., Куликов Г.А. Перцептивная значимость частотных и амплитудных характеристик гласных звуков при разной частоте основного тона // Доклады академии наук. 2009. Т. 429, № 3. С. 132–134.
7. Куликов Г.А., Андреева Н.Г. Перцептивно-значимые признаки гласных при разных формах их генерации // Сенсорные системы. 2009. Т. 23, № 3. С. 1–8.
8. Андреева Н.Г. Становление речи на ранних этапах онтогенеза // Нервная система. 2001. Вып. 35. С. 59–84.
9. Уплисова К.О. Акустический и аудиторский анализ гласноподобных звуков серого (*Psittacus erithacus*) и волнистого (*Melopsittacus undulatus*) попугаев // Сенсорные системы. 2006. Т. 20, № 3. С. 229–237.
10. Уплисова К.О. Говорящие птицы: особенности модифицированных гласных звуков // Сборник трудов XX сессии РАО. 2008. Т. 3. С. 110–113.
11. Цвикер Э., Фельдкеллер Р. Ухо как приемник информации / Пер. с нем. М.: Связь, 1965. 104 с.