

К ВОПРОСУ О ЗВУЧАНИИ РЕЧИ И ФОРТЕПИАНО¹

© 2024 г. В. А. Зверев^а, А. И. Малеханов^{а, *}

^аФедеральный исследовательский центр “Институт прикладной физики
им. А.В. Гапонова–Грехова Российской академии наук”,
ул. Ульянова 46, Нижний Новгород, 603950 Россия

*e-mail: almal@ipfran.ru

Поступила в редакцию 21.09.2023 г.

Принята к публикации 19.12.2023 г.

Рассмотрен вопрос слухового восприятия звуков речи и фортепиано, имеющих относительно слабый уровень основного тона на фоне высоких гармоник (формант голосовых связок и обертонов струны, соответственно). Показано, что слышимый спектр этих звуков является спектром огибающей интерференции колебаний соответствующих гармоник. Этот спектр содержит основной тон как доминирующий звук, который хорошо слышен в речи и в музыке, в то время как форманты и обертоны, изначально доминирующие в спектре звука речи и музыки, влияют преимущественно на тембр основного тона. На уровне гипотезы указано, что огибающая интерференции колебаний выделяется (детектируется) при этом в результате распространения звуковых волн в нелинейной среде жидкости улитки внутреннего уха.

Ключевые слова: слуховое восприятие, основной тон, обертоны, форманты, интерференция тональных сигналов, детектирование огибающей

DOI: 10.31857/S0320791924020151, EDN: YMPOUZ

ВВЕДЕНИЕ

В звуках речи и фортепиано отчетливо слышен звук основного тона, на который настроена голосовая связка или струна фортепиано соответственно. Однако в спектрах этих звуков основной тон часто не является преобладающим сигналом — его уровень может быть весьма низким сравнительно с уровнем обертонов (в звучании фортепиано) или формант (в звучании речи). Тем не менее, основной тон слышен нами как доминирующий звук, а обертоны и форманты влияют только на тембр звука основного тона [1]. В связи с этим возникает вопрос, который мы назовем *вопросом восприятия звука*: почему слушатель воспринимает основной тон в речи и в музыке как доминирующий звук, а обертоны и форманты, которые не просто присутствуют, но и доминируют в соответствующем спектре, влияют лишь на тембровую окраску основного тона?

Основной тон — это самая низкая по частоте мода собственных колебаний натянутой струны

и натянутых голосовых связок, закрепленных на концах. Наш повседневный опыт указывает на то, что при нажатии любой клавиши фортепиано или произнесении звука голосом, мы непременно слышим основной тон. К примеру, мы уверенно отличаем мужской голос от женского, прежде всего, по высоте основного тона.

В целом, поставленный вопрос восприятия звука относится к области психоакустики (см. например, [1–5]). В ряде работ [2, 3, 5] рассмотрен психоакустический механизм восприятия звука. Согласно этому механизму мы не слышим основной тон непосредственно, так как его уровень мал в спектре звука, но ощущаем его как сигнал разностной частоты соседних обертонов, равной частоте основного тона.

Подобная трактовка и сам механизм вызывают, однако, определенные сомнения. Ощущать можно только реально присутствующий сигнал, но нельзя ощущать тот сигнал, которого физически нет. Естественно предполагать, что музыкальный (речевой) сигнал основного тона должен присутствовать

¹ Это письмо оказалось последней публикацией в “Акустическом журнале” одного из самых продуктивных его авторов, члена-корреспондента РАН Виталия Анатольевича Зверева, ушедшего из жизни 6 марта 2024 г. В.А. Зверев является автором более 80 статей в “Акустическом журнале”, многие из которых могут быть отнесены к классическим работам современной акустики.

уже “на входе”, хотя он может быть при этом весьма слабым в сравнении с сигналами высоких гармоник – обертонов или формант.

Отмеченный психоакустический механизм опровергается не только этим общим рассуждением, но прямым опытом [3, стр. 8]. В этом опыте разность частот соседних обертонов путем прореживания спектра была увеличена вдвое, но слышимая частота основного тона при этом не изменилась.

Ответ на поставленный вопрос восприятия звука следует искать, по нашему мнению, непосредственно в области физической акустики, не привлекая для этого субъективные психоакустические факторы (не отвергая, разумеется, их роли вообще). Цель данной работы – попытаться дать такой ответ на основе физического описания процесса образования звуков фортепиано и звуков речи.

ОСНОВНОЙ ТОН, ОБЕРТОНЫ И ФОРМАНТЫ

Как известно, звуки фортепиано создаются колебаниями струн, а звуки речи – колебаниями голосовых связок. Как физический объект, голосовые связки можно представить как две струны, расположенные в глотке на пути потока воздуха из легких в полость рта. Начиная разговор или пение, человек натягивает голосовые связки, и в этом положении они возбуждаются проходящим через них потоком воздуха. Если звучание струны возникает после удара молоточка по ней в результате распространения колебаний, которые отражаются от закрепленных концов, то аналогичным образом действуют и голосовые связки, с той лишь разницей, что они возбуждаются переменным во времени потоком воздуха. Следовательно, оба этих процесса могут быть описаны с единых и хорошо известных позиций.

Как струны, так и связки обладают частотами собственных резонансных колебаний, геометрические формы которых называются модами. Самая низкочастотная мода есть основной тон, следующие за ней моды колебаний для струн и связок называются по-разному – обертонами и формантами, соответственно, их наличие играет существенную роль в решении вопроса восприятия звука. Удар молоточка по струне изначально возбуждает все моды свободных колебаний струны, а переменный поток воздуха через голосовые связки – все моды колебаний голосовых связок. Колебания струн фортепиано после их усиления резонансной деккой приводят к возбуждению звуковых волн в окружающем воздухе, их мы и слышим. В свою очередь, колебания голосовых связок модулируют проходящий через них воздушный поток, также возбуждая при этом звуковые волны. Эти волны усиливаются и корректируются по частоте резонансными

полостями голосового тракта человека, производя слышимый нами звук речи или пения.

Длина волны основного тона равна удвоенной длине струны или связки между точками их закрепления. Высшие моды обертонов и формант обладают длинами волн, кратно меньшими основной – между точками закрепления должно укладываться целое число длин их полуволн. На соответствующие частоты колебаний это условие распространяется только приблизительно, хотя и с хорошей точностью, поскольку скорость распространения волн вдоль струны или связки имеет слабую частотную зависимость.

Для примера, на рис. 1 показан спектр звука басовой ноты фортепиано, на рис. 2 (взят из сети Интернет) – спектр речи. Видно, что оба спектра состоят из набора дискретных частот, отвечающих модам основного тона и высшим модам, при этом уровень основной моды мал в сравнении с уровнем высших мод. На рис. 1 основной тон на частоте 27 Гц имеет уровень –9 дБ, в то время как уровень некоторых обертонов превышает 20 дБ, т.е. различие составляет ~30–40 дБ. Отметим, что это не всегда так и зависит от того, какая именно нота озвучена. Спектр звучания ноты может содержать основной тон с гораздо более высоким уровнем, но нет таких нот, основной тон которых заметно превалирует над обертонами. В спектре речи наблюдается качественно та же картина, как это видно по рис. 2.

Таким образом, мы убеждаемся, что поставленный выше вопрос восприятия звука имеет место. Попытаемся теперь ответить на него.

ВАРИАНТ ОТВЕТА НА ВОПРОС ВОСПРИЯТИЯ ЗВУКА

Итак, опираясь на изложенные выше известные положения, рассмотрим, каким образом можно дать непротиворечивый ответ на поставленный вопрос.

Выполним следующий численный опыт. Широкополосный звуковой сигнал, спектр которого показан на рис. 1 и в котором, что важно, практически нет явно выраженного основного тона, подвергнем обычному детектированию. Результат показан на рис. 3. Видно, что уровень основного тона на выходе детектора существенно вырос и достигает уровня обертонов. Далее, если в спектре на рис. 3 убрать высокие частоты, что стандартно делается в результате детектирования входного сигнала в радиоприемнике, то основной тон станет превалирующим. Этому сильно поможет то, что спектр, который мы слышим, обладает логарифмическим масштабом по частоте и определяется в октавах. Назовем такой спектр октавным спектром, и он имеет существенные отличия от обычного Фурье-спектра.

В работе [4] спектр звука, формируемый фильтрами базилярной мембраны внутреннего уха [3], получен обоими способами: как обычный Фурье-спектр и как октавный спектр. В обычном спектре разрешение спектральных компонент по частоте определяется соотношением, которое не имеет зависимости от частоты:

$$\Delta f = \frac{1}{T}, \tag{1}$$

где T – заданное время интегрирования. В октавном спектре время интегрирования обратно пропорционально текущей частоте f :

$$T = \frac{r}{f}, \tag{2}$$

где r – величина, зависящая от качества слуха человека, и она всегда больше 12 [4]. Согласно (2), в октавном спектре разрешение по частоте определяется как

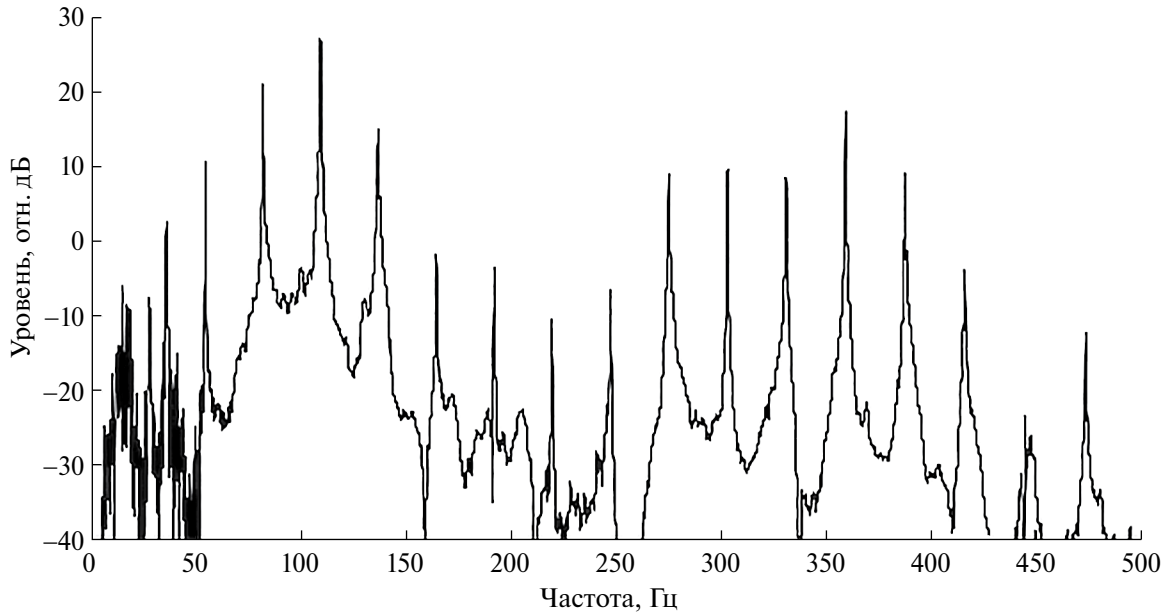


Рис. 1. Спектр звука фортепиано на примере ноты Ля субконтравты.

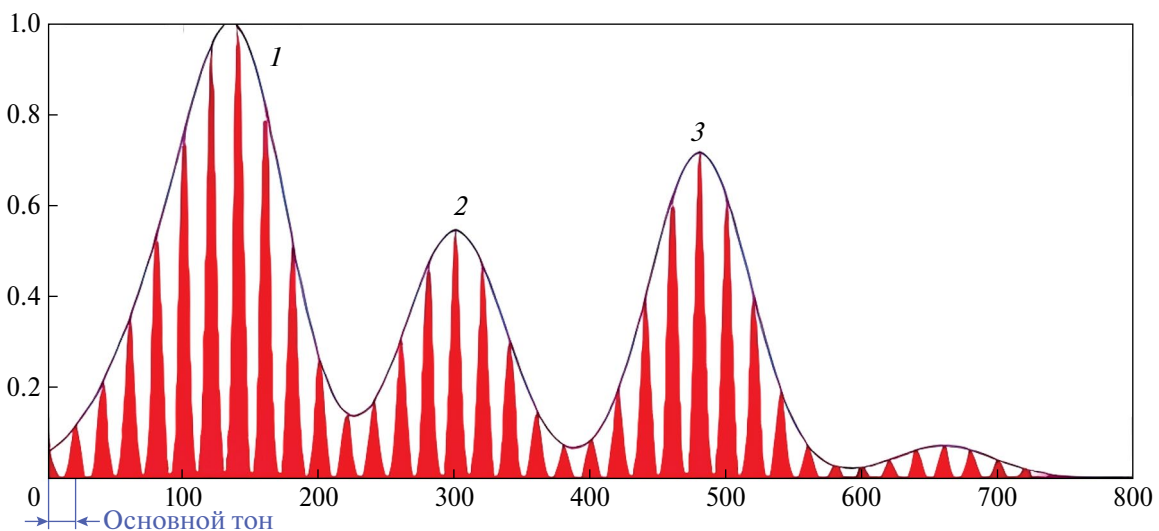


Рис. 2. Спектр речи (качественное изображение).

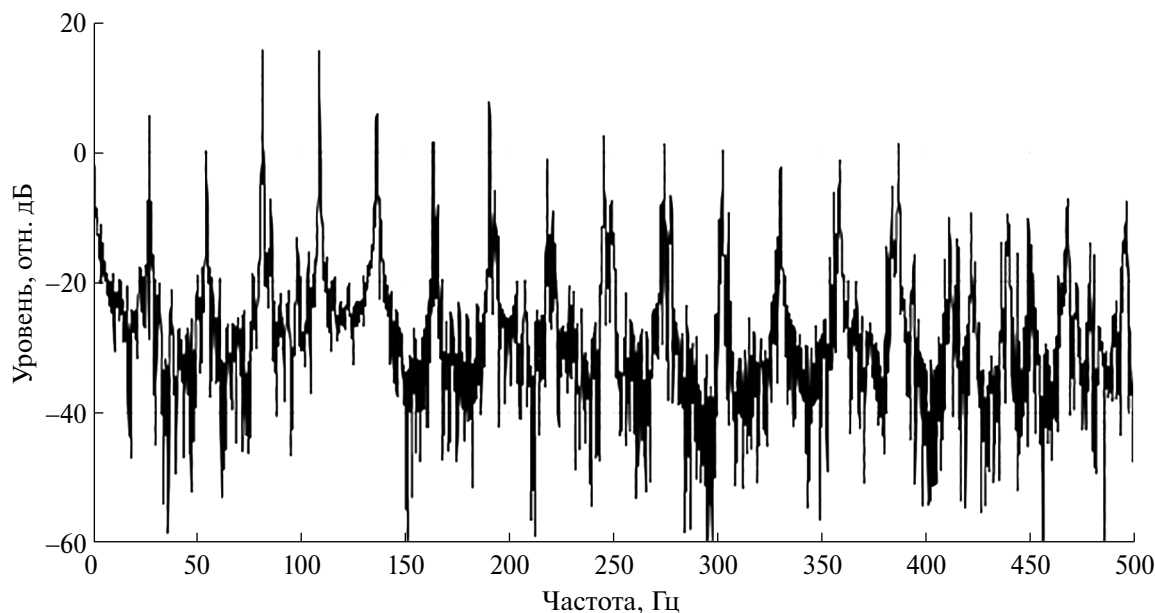


Рис. 3. Спектр звука басовой ноты (рис. 1) после линейного детектирования.

$$\Delta f = \frac{f}{r}, \quad (3)$$

Уровень составляющих в обоих спектрах определяется отношением частоты f к разрешению по частоте:

$$U = \frac{f}{\Delta f}, \quad (4)$$

Поскольку в обычном спектре разрешение по частоте не зависит от самой частоты, то уровень (4) спектральных составляющих в нем растет пропорционально частоте. Но в характерном для слухового восприятия октавном спектре ситуация совсем другая – уровень спектральных составляющих не растет с частотой и остается постоянным. В результате получается, что уровень спектральных составляющих октавного спектра для одного и того же входного сигнала убывает пропорционально частоте с декрементом 6 дБ на октаву.

Итак, основной тон в октавном спектре с учетом фильтрации детектирования и ослабления уровня высоких частот становится преобладающим звуком, несмотря на его слабый уровень в исходном (входном) сигнале. Остается уточнить, что именно осуществляет детектор внутреннего уха для того, чтобы обеспечить этот результат.

Проведем математическое моделирование нужных нам преобразований звукового сигнала на примере суперпозиции эквидистантных по частоте обертонов одного уровня с частотами 100, 120, 140, 160 Гц (рис. 4). Разностная частота

20 Гц играет здесь роль основного тона, которого изначально нет среди указанных сигналов. На рис. 4а показана временная форма исходного суммарного сигнала, где хорошо видна ожидаемая периодическая последовательность максимумов и минимумов (интерференционные биения), на рис. 4б – временная зависимость уровня сигнала (его огибающая). На рис. 4в показан спектр исходного сигнала, который содержит только заданные четыре компонента и более ничего (очевидно), на рис. 4г – спектр интерференционного сигнала после детектирования. На этом рисунке хорошо видно появление основного тона (20 Гц), при этом его уровень сопоставим с уровнем гармоник (обертонов).

Таким образом, физический механизм хорошей слышимости основного тона может быть основан на таком преобразовании сложного сигнала, исходно содержащего значительный набор интенсивных гармоник, в результате которого осуществляется детектирование зависимости уровня интерференции гармоник от времени. Важно подчеркнуть, что саму огибающую мы не слышим непосредственно, так как ее частота – частота основного тона в нашем модельном примере (см. рис. 4г) – много больше 7 Гц, которая есть предельная (сверху) частота слышимости биений близких звуковых частот [6].

Мы предполагаем, что нужная нам операция выделения огибающей входного сигнала может быть осуществлена при прохождении звуковых волн через нелинейную среду внутреннего уха наподобие того, как работает параметрическая

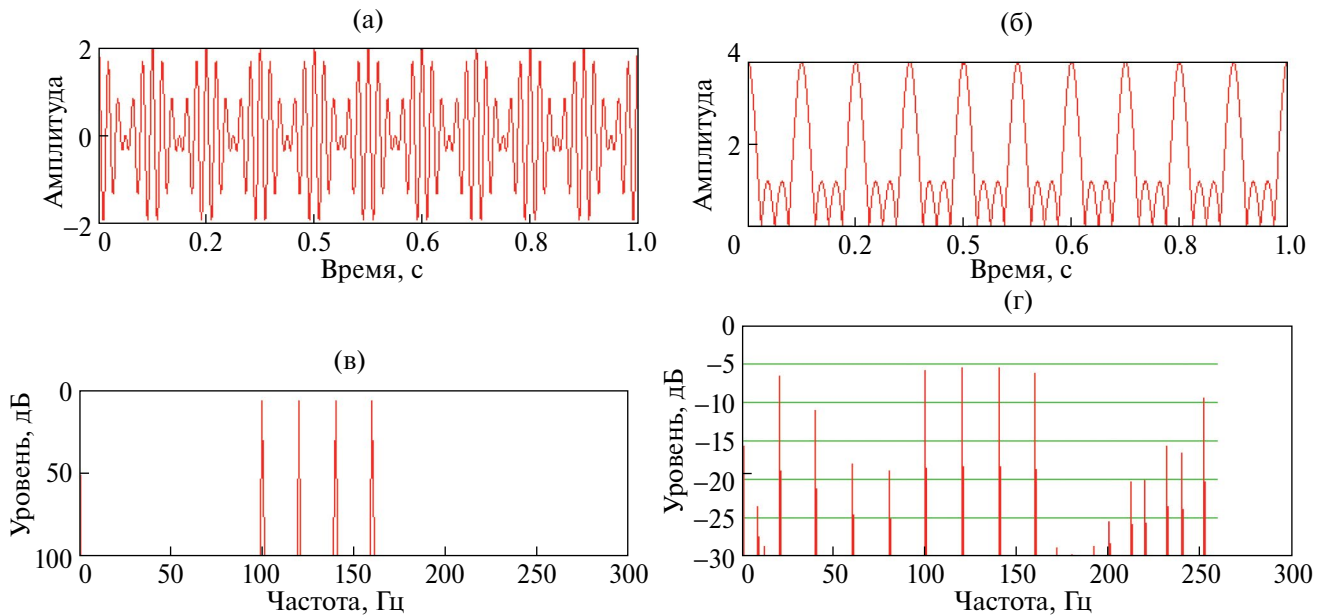


Рис. 4. Интерференция четырех эквидистантных гармоник на частотах 100, 120, 140, 160 Гц: (а) – суммарный сигнал и (б) – его огибающая, (в) – спектр суммарного сигнала, (г) – спектр огибающей.

антенна в гидроакустике [7, 8]. Излучаемые параметрической антенной два интенсивных высокочастотных сигнала соответствуют в нашей постановке двум обертонам, изначально отсутствующий сигнал разностной частоты – основному тону, который формируется (детектируется) в объеме среды распространения исходных сигналов. В составе слуха “кандидат” на такую нелинейную среду есть, и он представляется единственным – это жидкость внутреннего уха.

Данный “параметрический” механизм обсуждается здесь на уровне гипотезы, поскольку обоснование его требует отдельного исследования с учетом конкретных значений физических параметров аппарата внутреннего уха, что выходит за рамки данной работы. Можно только предполагать, что эти параметры способны обеспечить необходимый эффект детектирования. На это объективно указывает наш повседневный опыт – явная слышимость основного тона.

В заключение, также в качестве гипотезы, приведем возможное описание действия слухового детектора в том случае, когда входной звуковой сигнал не содержит обертонов, т.е. когда отсутствуют сложный сигнал интерференции обертонов и его низкочастотная огибающая. Тогда, как мы предполагаем, отсутствует и фильтрация низких частот после детектора: сигналы, не содержащие огибающей интерференции, после детектирования фильтрации не подвергаются. Поэтому во всех случаях отсутствия в звуковом сигнале обертонов мы слышим его истинный спектр.

ВЫВОДЫ

Подводя итог нашему рассмотрению, мы делаем вывод, что основной тон в спектре звучания музыкальных нот фортепиано или человеческого голоса образуется, становясь доминирующим, благодаря детектированию огибающей звукового сигнала и подавлению его высоких гармоник при формировании октавного спектра базиллярной мембраной. Детектором, выделяющим огибающую интерференции обертонов (в случае звуков фортепиано) или формант (в случае голосовых сигналов), может являться нелинейная среда жидкости улитки внутреннего уха, в которой распространяются звуковые волны, поступающие на вход ушной раковины.

Авторы благодарны профессору Л.А. Островскому за указание на аналогию с работой параметрической антенны и ту роль, которую играет жидкость внутреннего уха в качестве нелинейной среды, В.Н. Голубеву и И.Н. Диденкулову (ИПФ РАН) за помощь в работе и обсуждения.

СПИСОК ЛИТЕРАТУРЫ

1. Вологдин Э.И. Слух и восприятие звука. Санкт-Петербург, 2012.
2. Галембо А.С. Фортепиано. Качество звучания. М.: Легпромбытиздат, 1987. 168 с.
3. Алдошина И.А. Основы психоакустики // Звукорежиссер. 1999.

4. *Зверев В.А.* Влияние темпа исполнения на слышимость басовых нот // Акуст. журн. 2021. Т. 67. № 3. С. 338–343.
5. *Кравчун П.Н.* Рецензия на статью В.А. Зверева // Акуст. журн. 2021. Т. 67. № 3. С. 343–344.
6. Интерференция и биения в физике и в музыке // <https://www.audiomania.ru/content/art-4081.html>
7. *Зверев В.А.* Как зародилась идея параметрической антенны // Акуст. журн. 1999. Т. 45. № 5. С. 685–692.
8. *Есипов И.Б., Зверев В.А., Калачев А.И., Наугольных К.А.* О низкочастотном параметрическом излучении звука // Акуст. журн. 1976. Т. 22. № 4. С. 606–608.

On the Question About the Sound of Speech and Piano

V. A. Zverev¹, A. I. Malekhanov^{1, *}

¹*Federal Research Center “Institute of Applied Physics named after. A.V. Gaponov–Grekhov
Russian Academy of Sciences”,
st. Ulyanova 46, Nizhny Novgorod, 603950 Russia*

**e-mail: almal@ipfran.ru*

The issue of auditory perception of speech and piano sounds, which have a relatively weak fundamental tone level against the background of high harmonics (formants of the vocal cords and string overtones, respectively), is considered. It is shown that the audible spectrum of these sounds is the spectrum of the interference envelope of oscillations of the corresponding harmonics. This spectrum contains the fundamental tone as the dominant sound, which is clearly heard in speech and music, while the formants and overtones, initially dominant in the sound spectrum of speech and music, primarily affect the timbre of the fundamental tone. At the hypothesis level, it is indicated that the vibration interference envelope is isolated (detected) as a result of the propagation of sound waves in the nonlinear fluid medium of the cochlea of the inner ear.

Keywords: auditory perception, fundamental tone, overtones, formants, interference of tonal signals, envelope detection