

С помощью описанного метода была произведена сегментация около 150 звуко-сочетаний ГГ, СС, СГ (без взрывных и аффрикат), произнесенных тремя дикторами. Надежность сегментации по трем дикторам составила 98%. Данный метод, как и в [3], может использоваться для выделения первичных сегментов в слове, причем надежность сегментации в обоих случаях существенно зависит от величины кадра анализа. Экспериментально было установлено, что эта величина может выбираться в пределах 20–30 мсек. При уменьшении или увеличении кадра анализа от указанных временных интервалов наблюдается менее выраженная декорреляция векторов  $s$  и  $s'$  на переходных участках, что затрудняет работу алгоритма выделения локальных минимумов сегментирующей функции и понижает в целом надежность сегментации (в данной работе величина кадра составляла 23 мсек).

Сравнительный анализ двух методов (с использованием функций поперечного сечения и спектральных срезов [3]) показал, что время сегментации с помощью описанного метода сокращается в 7–8 раз, так как число координат в векторах  $s$  и  $s'$  составляет всего 12–14 вместо 100. Кроме того, переходные участки отмечаются более выраженными локальными минимумами, чем в работе [3]. Однако при сегментации фонем, очень близких по месту артикуляции, локальный минимум сегментирующей функции может не достигнуть пороговой константы, что приведет к выделению одного сегмента вместо двух. В таких случаях могут оказаться полезными другие методы сегментации, не критичные по отношению к близости места артикуляции смежных фонем.

В заключение автор благодарит Б. М. Лобанова за полезное обсуждение работы и критические замечания.

#### ЛИТЕРАТУРА

1. Распознавание слуховых образов, под ред. Загоруйко Н. Г., Волошина Г. Я. Новосибирск, «Наука», 1970.
2. Чистович Л. А., Кожевников В. А., Алякринский В. В., Бондарко Л. В., Голузина А. Г., Клаас Ю. А., Кузьмин Ю. И., Лисенко Д. М., Люблинская В. В., Федорова Н. А., Шупляков В. С., Шуплякова Р. М. Речь. Артикуляция и восприятие. М.—Л., «Наука», 1965.
3. Деркач М. Ф., Гумецкий Р. Я., Мишин Л., Оверченко М., Чабан М. Восприятие речи в распознающих моделях. Изд-во Львовского ун-та, 1971.
4. Скучик Е. Основы акустики, т. II. М., Изд-во иностр. лит., 1959.
5. Wakita H. Direkt Estimation of the Vocal Tract Shape by Inverse Filtering of Acoustic Speech Waveforms. IEEE Transactions on Audio and Electroacoustics, 1973, AU-21, 5, 417–427.
6. Markel J., Gray A. On Autocorrelations Equations as Applied to Speech Analysis. IEEE Transactions on Audio and Electroacoustics, 1973, AU-21, 2, 69–79.

Минский радиотехнический институт

Поступила  
17 февраля 1978 г.  
После повторного исправления  
26 сентября 1978 г.

УДК 534.684:534.42

#### РОЛЬ ОБУЧЕНИЯ ПРИ ВОСПРИЯТИИ ИСКАЖЕННОЙ РЕЧИ

М. А. Сапожков

Приведены данные по распознаванию речи слухом в случае амплитудно-частотных искажений ее. Определена роль фактора тренировки к искажениям речи и показано, что при длительной тренировке и обучении слушатель может хорошо воспринимать сильно искаженную речь.

Несмотря на ряд работ [1], до сих пор некоторые исследователи сомневаются, что формантные картины речи (взаимное расположение формант и антиформант и форма формантных переходов) являются определяющим признаком для распознавания звуков речи слухом человека [2]. Чтобы рассеять эти сомнения, были проведены эксперименты и было установлено, что при сильном подчеркивании спектральных составляющих речи около 300 гц звук «о» воспринимается как звук «у», а при подчеркивании их около 500 гц звук «у» воспринимается как звук «о», при подчеркивании спектральных составляющих в области 7000 гц звук «ш» воспринимается как звук «с», а при подчеркивании их в области около 4000 гц звук «с» воспринимается как звук «ш». При этих экспериментах слушатели не знали, что передают, и могли узнавать передаваемые слова или бессмысленные слоги или по смыслу (для

слов), или по своему слуховому ощущению (для слогов). Поэтому при передаче бессмысленных слогов большинство слушателей воспринимало слоги так, как они звучали для них, а поэтому при искажениях распознавали звуки неправильно. При восприятии слов слушатели корректировали свое распознавание по смыслу, но если при искажениях звуков речи слова изменяли свой смысл, то слушатели не могли догадаться, какое слово передавалось, и тоже распознавали, как правило, неправильно. Например, слово «мол» воспринимали как «мул», и наоборот; слово «мус» воспринимали как «муш», и наоборот. Это свидетельство о распознавании звуков речи по формантным картинам при невозможности коррекции правильности приема. При этих экспериментах было установлено, что большую роль в распознавании звуков речи при наличии их искажений имеет тренировка к приему искаженной речи в форме обучения слушателей.

Для определения роли фактора обучения слушателей был поставлен ряд экспериментов. Эксперименты [3] были проведены как с вещательными трактами передачи речи, так и синтезаторами речи. Поскольку были возражения против применения синтезаторов речи как устройств, в которых могут нарушаться привычные связи между звуками речи, для данных экспериментов были использованы тракты для передачи речи типа как вещательных, так и телефонных. В вещательный тракт вводились искажения типа подчеркивания определенных участков частотного диапазона (в областях 300—500, 1000 и 4000—7000 *гц*). Для этого использовались резонансные контуры с добротностью около 10. В телефонные тракты вводилось резкое ограничение частотного диапазона снизу (на частоте 500 *гц*) и сверху (на частоте 1800 *гц*) или применялась инверсия спектра речи с помощью устройств многоканальной связи.

В качестве испытательного материала были применены артикуляционные слоговые таблицы по пяти для каждого цикла измерений, т. е. по 250 слогов; 40 слушателей были разбиты на четыре группы. Для первой группы при передаче таблиц каждый слог повторяли трижды: сначала — неискаженный, затем дважды — искаженный. После передачи пяти таблиц их повторяли только с искажением и слушателям предлагали записать их не так, как они звучат, а их неискаженное произнесение. Подобная процедура повторялась несколько раз, причем каждый раз порядок чтения слогов в таблицах изменялся. Повторение проводилось до тех пор, пока не достигались установившиеся значения разборчивости слогов. Вторая группа слушателей прослушивала те же таблицы, но без передачи неискаженных слогов, с однократным их произнесением. Эти слушатели проверяли правильность записи слогов, как и полагается при артикуляционных испытаниях, т. е. после передачи цикла таблиц. Третья группа слушателей прослушивала те же таблицы, причем перед каждым испытательным слогом произносилась фраза из 3—4 слов, например, «Что означает термин...», «Какое слово будет произнесено...». Эти слушатели не проверяли правильность записей слогов (это выполняли другие лица), поэтому их тренировка ограничивалась прослушиванием вводных фраз. Наконец, четвертая группа слушателей только прослушивала те же таблицы и в своих записях целиком должна была полагаться на свой слух, т. е. никакой тренировки у нее не было.

Для каждого вида искажений, кроме инверсии спектра, были определены предельно-достижимые значения разборчивости слогов для всех четырех групп слушателей. Для первых двух групп слушателей определялось также необходимое число циклов обучения для достижения установившегося значения разборчивости слогов. В таблице приводятся полученные результаты измерений.

Анализ этих данных показывает, что даже для самых сильных искажений (инверсия спектра) может быть достигнута хорошая разборчивость речи, но для этого требуется длительное обучение приему речи при таких искажениях. Оказалось, что

Тип искажений	Обученные слушатели		Артикулянты		Необученные слушатели	
	разборчивость, %	число циклов	разборчивость, %	число циклов	с вводной фразой	без вводной фразы
					разборчивость, %	
Подчеркивание	96	3	82	5	68	63
Ограничение частотного диапазона	90	5	75	12	43	41
Инверсия спектра	58*	25	—**	—**	—**	—**

\* Это значение еще не было предельным, так как процент правильно принимаемых слогов очень медленно, но неуклонно повышался и, возможно, при очень большом числе циклов мог достичь несколько большего значения.

\*\* Инверсный спектр этим группам не передавали, так как было заранее известно, что они не поймут передаваемых слогов.

при амплитудно-частотных искажениях разборчивость слогов после обучения достигает почти таких же значений, как и при приеме неискаженных слогов, и для этого достаточна небольшая тренировка к приему с такими искажениями. Совершенно необученные слушатели (четвертая группа) даже после многократного прослушивания искаженных слогов повышают количество правильно принятых слогов незначительно, и величина разборчивости слогов получается для них меньше, чем это следует при расчете разборчивости речи. Применение вводной фразы для тренировки почти не изменяет данных по сравнению с предыдущими данными, поэтому разборчивость слогов для третьей группы получилась почти такая же, как и для четвертой. Разборчивость слогов для артикулянтов (вторая группа) получилась гораздо выше, чем для необученных слушателей. Это объясняется тем, что артикулянты узнавали свои ошибки при проверке записей. Полученные для этой группы результаты согласуются с теорией разборчивости речи, что вполне понятно, так как в основу расчетных соотношений для артикуляционного метода положена такая проверка таблиц.

Таким образом, хотя восприятие звуков речи происходит по формантным картинкам, но при соответствующем обучении слушателей возможен прием речи с получением высокой разборчивости даже при значительных искажениях. Подобный метод равносильен обучению человека приему иностранной речи с переводом ее на свой язык.

#### ЛИТЕРАТУРА

1. Фланаган Дж. Л. Анализ, синтез и восприятие речи. Пер. с англ. под ред. А. А. Пирогова. М., «Связь», 1968.
2. Вокодерная телефония, под ред. А. А. Пирогова. М., «Связь», 1974.
3. Сапожков М. А. Заметность амплитудно-частотных искажений речи. Электросвязь, 1974, 6, 23—25.

Московский электротехнический институт связи

Поступила  
2 июня 1978 г.

УДК 534.26

### О ВЛИЯНИИ СДВИГА, ИНЕРЦИИ ВРАЩЕНИЯ И ПРОДОЛЬНЫХ КОЛЕБАНИЙ НА ИЗЛУЧЕНИЕ ЗВУКА ПЛАСТИНОЙ ПРИ ВОЗБУЖДЕНИИ ЕЕ СОСРЕДОТОЧЕННОЙ СИЛОЙ

Ю. Д. Сергеев

В опубликованных к настоящему времени работах, посвященных исследованию излучения звука неограниченными упругими пластинами, их поперечные колебания описываются, как правило, классическим уравнением изгибных волн. Однако на частотах, близких к резонансу совпадения, точность этого уравнения недостаточна [1]. Кроме того, и продольные волны в пластине [2] могут оказать определенное влияние на излучение ею звука. Поэтому была рассмотрена задача об излучении звука возбуждаемой сосредоточенной поперечной гармонической силой упругой неограниченной пластиной в более точной постановке с использованием известных из работ [3, 4] уравнений изгибных и продольных колебаний. Пусть пластина одной стороной граничит с жидким полупространством. Соответствующая система уравнений имеет вид

$$(1) \quad \left( \Delta + \frac{\rho_{\text{п}} \omega^2}{\sigma_{\text{п}}} \right) \left( \mathcal{D} \Delta + \frac{\rho_{\text{п}} h^3}{12} \omega^2 \right) w - \omega^2 \rho_{\text{п}} h w = \\ = \left( 1 - \frac{\mathcal{D}}{\mu \sigma h} \Delta - \frac{\rho_{\text{п}} h^2 \omega^2}{12 \sigma_{\text{п}}} \right) (F_0 \delta(x) - p|_{z=0}),$$

$$(2) \quad c_1 \Delta v + \omega^2 v = \frac{h}{4E_1} [(1 - \sigma_1^2) \omega^2 - c_1^2 \Delta] (F_0 \delta(x) + p|_{z=0}),$$

$$(3) \quad j\omega(w - v) = \zeta|_{z=0},$$

где  $p$ ,  $\zeta$  — звуковое давление и колебательная скорость в среде,  $w$ ,  $v$  — компоненты поперечного смещения пластины при изгибных и продольных колебаниях;  $\mathcal{D}$ ,  $\mu$ ,  $\rho_{\text{п}}$ ,  $\nu$ ,  $h$ ,  $c_1$  — изгибная жесткость, модуль сдвига, плотность, коэффициент Пуассона,