

ИДЕНТИФИКАЦИЯ ДИКТОРА ПО ЧАСТОТАМ ФОРМАНТ, ИЗМЕРЕННЫМ СИНХРОННО С ОСНОВНЫМ ТОНОМ

ВАЛЕРИЙ ГИТЛИН

Кафедра "Вычислительная техника"  
Устиновский механический институт  
Устинов, Удмуртия, СССР, 426000

Идентификация диктора по голосу есть процесс идентификации некоторой физической системы, который требует достаточно точного определения параметров системы. В качестве таких параметров могут быть взяты частоты формант и их траектории. Однако процесс выделения формант достаточно труден. Эти трудности объясняются влиянием голосового источника на форму речевого сигнала и влиянием аппаратуры анализа. Учет этих влияний позволил снизить ошибку идентификации 104 дикторов мужчин по траекториям формант парольной фразы с 12,8% до 5,3%.

ВВЕДЕНИЕ

Человека можно представить как некоторую физическую систему. Речевой сигнал есть продукт этой системы и отражает ее конкретные физические особенности. В качестве меры параметров речевого тракта можно использовать частоты формант [1].

Структурная схема выделителя периода форманты показана на фиг. 1. Речевой сигнал разделялся формантными фильтрами  $FF_n$  ( $n = 1, 2, 3$ ) на формантные полосы. В каждой полосе синхронно с основным тоном измерялся период форманты по методу Кампанеллы и Коултера [2]. Основной тон выделялся пиковым методом по речевому сигналу, ограниченному полосой 4,5 кГц [3].

В разделе I настоящей работы выполнен анализ погрешностей выделения частоты формант. В разделе II представлена методика оптимизации формантных фильтров. Результаты экспериментов по идентификации диктора даны в разделе III, раздел IV - заключение.

I. ПОГРЕШНОСТИ ВЫДЕЛЕНИЯ ЧАСТОТ ФОРМАНТ

A. Влияние голосового источника

Был выполнен анализ реального речевого сигнала. Вычислялась огибающая спектра на интервале открытых (ОС) и закрытых (ЗС) голосовых связок и на периоде основного тона (ОТ) [4] при помощи методики линейного предсказания. Интервалы ОС, ЗС и ОТ выделялись ручным способом по осциллограммам речи. Огибающие спектров для гласной / i / из слова "электричество" показаны на фиг. 2 [5]. Аналогичные результаты получены Ларером и др. [6].

Из фиг. 2 видно, что в реальной речи открывание голосовых связок изменяет частоту и ширину формант, причем меньше всего подвержена изменениям резонансная частота первой форманты  $F_{m1} = [F_1^2 - (B_1/2)^2]^{1/2}$ . Спектр сигнала, вычисленный на периоде ОТ, был ближе к спектру, вычисленному на интервале ЗС, а не на интервале ОС.

B. Оценка взаимного влияния формант

Пусть сигнал искомой форманты записывается как  $U_i e^{-\sigma_i t} \sin(\omega_i t + \varphi_i)$ , а сигнал

мешающей форманты -  $U_2 e^{-\sigma_2 t} \sin(\omega_2 t + \varphi_2)$ . Для того, чтобы средняя частота пересечений нуля соответствовала измеряемой форманте, необходимо выполнить условие [9,5]:

$$U_1 > U_2 e^{-(\sigma_2 - \sigma_1)t} \mid \omega_1 > \omega_2, \quad (1)$$

$$U_1 > U_2 \frac{\omega_2}{\omega_1} e^{-(\sigma_2 - \sigma_1)t} \mid \omega_1 < \omega_2.$$

Из формулы (1) видно, что момент максимального подавления мешающей форманты зависит от амплитудных соотношений формант, частот формант и величин их затуханий.

C. Собственные колебания формантного фильтра

Представим речевой сигнал как [1]:

$$u(t) = \sum_{n=1}^N (-1)^n U_n e^{-\sigma_n t} \sin(\omega_n t + \varphi_n). \quad (2)$$

Если сигнал  $u(t)$ , определяемый формулой (2) подать на формантные фильтры  $FF_n$  (фиг.1), то в каждом фильтре будут возникать собственные колебания с каждым новым возбуждением затухающей синусоиды [7], т.е. с каждым новым возбуждением речевого тракта.

Пусть  $FF$  составлен из  $L$  резонансных звеньев, включенных последовательно. Передаточная характеристика звена:

$$K_\ell(p) = -G_\ell (p + 2\alpha_\ell) / [(p + \alpha_\ell)^2 + \omega_{0\ell}^2], \quad (3)$$

где  $\ell = 1, 2, \dots, L$  - номер звена,  $G_\ell$  - коэффициент передачи,  $\alpha_\ell$  - затухание,  $\omega_{0\ell} = (\omega_{0\ell}^2 - \alpha_\ell^2)^{1/2}$  - резонансная частота звена с учетом потерь.

Подадим на  $FF_n$  с передаточной характеристикой  $K(p) = \prod K_\ell(p)$ , сигнал вида (2). Выходной сигнал фильтра можно представить как [5]:

$$j \dot{u}^{out}(t) = U_i^{out} e^{-\sigma_i t + j(\omega_i t + \beta_i)} \bar{N}_i(t), \quad (4)$$

где

$$\bar{N}_i(t) = 1 + \sum_{n=1, n \neq i}^N \frac{U_n^{out}}{U_i^{out}} e^{-(\sigma_n - \sigma_i)t + j[(\omega_n - \omega_i)t + \beta_n - \beta_i]} + \sum_{\ell=1}^L \frac{U_{n\ell}^{out}}{U_i^{out}} e^{-(\alpha_\ell - \sigma_i)t + j[(\omega_{0\ell} - \omega_i)t + \gamma_{n\ell} - \beta_i]} \quad (5)$$

- комплексный поправочный коэффициент, оценивающий влияние мешающих факторов на форму сигнала  $i$ -ой форманты,  $U_n^{out}$ ,  $\beta_n$  - амплитуда и фаза вынужденного колебания на выходе  $FF_n$ ;  $U_{n\ell}^{out}$ ,  $\gamma_{n\ell}$  - амплитуда и фаза свободного колебания в  $\ell$ -ом звене  $FF_n$ . Формулы для расчета  $U_n^{out}$ ,  $\beta_n$ ,  $U_{n\ell}^{out}$  и  $\gamma_{n\ell}$  даны в работе [8].

Примерный вид графика  $|\bar{N}_i(t)|$  для  $\sigma_n < \sigma_i$  показан на фиг. 3, где  $\delta$  - допустимое отклонение  $|\bar{N}_i(t)|$  от единицы,  $T_d$  - время задержки измерений,  $T_m$  - интервал времени, в течение которого возможны измерения параметров формант. Конкретный вид функции  $|\bar{N}_i(t)|$  зависит от соотношения значений  $\alpha_\ell$ ,  $\sigma_n$ ,  $\omega_{0\ell}$ ,  $\omega_n$ .

Анализ выражения (5) выполнялся путем численного моделирования на ЭВМ. Как показывают расчеты, если  $\alpha_\ell \geq 3\sigma$ , то скорость уменьшения  $T_d$  становится незначительной. Поэтому брать  $\alpha_\ell > 3\sigma$  нецелесообразно.

Можно показать [7], что при увеличении числа звеньев фильтра, скорость нарастания  $|\bar{N}_i(t)|$  снижается, а  $T_d$  - растет. Для  $\omega_{0\ell} = \omega_{0i}$ ,  $\ell = 1, 2, \dots, L$ ;  $\alpha_\ell = \alpha_0$  для  $\ell = 1, 2, \dots, L$  и  $L = 3$  имеем [7]:

$$\alpha_0 T_d \cong 6. \quad (6)$$

$FF_n$  должен обеспечивать малое значение  $T_d$ , большое  $T_m$  и высокую избирательность по частоте. Эти требования противоречивы и необходима процедура оптимизации для получения наилучших характеристик фильтра.

II. ВЫБОР ПАРАМЕТРОВ ФОРМАНТНОГО ФИЛЬТРА

Было решено [8] остановиться на трехзвенной схеме формантного фильтра ( $L = 3$ ). Если выбрать  $T_d < T_{min} = 2$  мс, то из формулы (6) получаем, что  $\alpha \geq 3000$  и полоса пропускания отдельного звена  $FF_n$   $\Delta f_n \geq \alpha/\pi = 955$  Гц.

Оптимизация параметров  $FF_n$  выполнялась [9] путем поиска минимума функции  $Q(i, n, j) = U_{nj}^{out} / U_{ij}^{out}$ , (7)

где  $j = 1, 2, \dots, 6$  - индекс фонемы, для которой вычисляется функция  $Q$ ,  $i = 1, 2, 3$  - индекс искомой форманты,  $n = 1, 2, 3, 4$  - индекс мешающей форманты. При вычислении  $Q$  полагалось, что амплитуды формант одинаковы. С целью повышения избирательности диапазон  $F_2$  был разбит на два поддиапазона. Параметры рассчитанных  $FF_n$  представлены в таблице, а их частотные характеристики показаны на фиг. 4.

Таблица

Форманта	Звено	Резонансная частота, Гц	Полоса, Гц
$F_1$	1	170	720
	2	240	740
	3	280	740
$F_{2-1}$	1	950	800
	2	1045	800
	3	1215	800
$F_{2-2}$	1	1235	1125
	2	1365	800
	3	1770	800
$F_3$	1	2410	720
	2	2570	925
	3	3260	1045

### III. ЭКСПЕРИМЕНТЫ ПО ИДЕНТИФИКАЦИИ ДИКТОРА

На этапе предварительных экспериментов дикторы идентифицировались по средним частотам формант, выделенных на стационарных участках гласных. В качестве речевого материала использовались шесть русских гласных (/а/, /о/, /у/, /э/, /и/ и /ы/), произнесенных в составе слов семью дикторами (три женщины, четверо мужчин). В условиях машинного зала было сделано три сеанса записей по семь произнесений каждой фонемы в одном сеансе. Шесть произнесений использовались для обучения, седьмое - как контрольное. Результаты идентификации усреднялись по трем сеансам записи. В качестве меры сходства использовалось Эвклидово расстояние между частотами формант эталонной и контрольной выборки. В качестве формантных фильтров использовались либо рассчитанные

формантные фильтры (РФФ), либо фильтры, параметры которых выбирались эмпирически (ЭФФ) [3].

Число ошибок идентификации мужчин составило 19% и 44%, женщин - 27% и 50% при использовании РФФ и ЭФФ соответственно. РФФ обеспечили существенное снижение количества ошибок идентификации.

На этапе основных экспериментов дикторы идентифицировались по траекториям формант парольной фразы: "С[ода в луже убывала] слабо". В экспериментах участвовало 104 диктора мужчины. Фраза записывалась два раза в одном сеансе испытаний в условиях машинного зала. Первое произнесение использовалось для обучения, второе - как контрольное. Для описания траекторий формант применялся алгоритм Ламиса [10]. В качестве меры сходства использовалось Эвклидово расстояние между коэффициентами разложения траекторий по ортогональным полиномам для трех формант плюс коэффициент корреляции между эталонной и контрольной выборками.

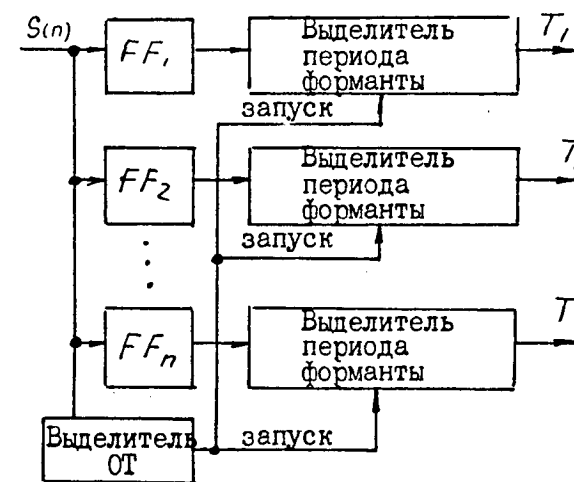
Ошибка идентификации составила для РФФ 5,3%, а для ЭФФ - 12,8%. РФФ и в этом случае обеспечил уменьшение количества ошибок идентификации.

### IV. ЗАКЛЮЧЕНИЕ

Снизить влияние источника и аппаратуры анализа на точность измерения параметров формант и, тем самым, повысить надежность идентификации диктора можно путем анализа временной функции речевого сигнала синхронно с основным тоном. Параметры формантного фильтра должны быть выбраны оптимальным образом с наиболее короткой импульсной реакцией и с максимальной возможной избирательностью по частоте. Траектории формант обеспечивают повышение надежности идентификации диктора по сравнению со средними значениями частот формант стационарных участков гласных.

### ЛИТЕРАТУРА

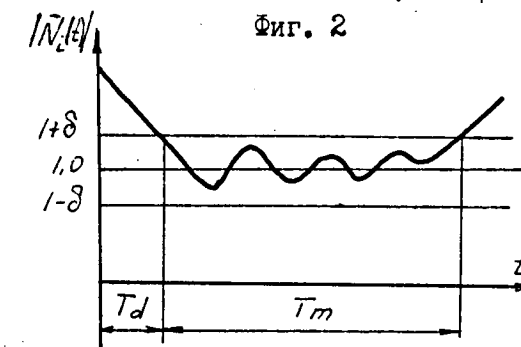
1. Фант Г. Акустическая теория речеобразования. - М.: Наука, 1964. - С. 284.
2. Campanella S.J., Coulter D.C. Formant Period Tracker. Pat. USA. № 3335225
3. Сапожков М.А. Речевой сигнал в кибернетике и связи. - М.: Связь, 1963. - С. 422.
4. Гитлин В.Б., Сметанин А.М. О повышении точности измерения параметров формант // Проблемы построения систем понимания речи. - М.: Наука, 1980. - С. 109-115.
5. Гитлин В.Б. и др. Выбор интервалов измерений частоты и ширины формант. Тез. докл. АРСО-Х. - Тбилиси: Мецниереба, 1978. - С. 20-22.
6. Larez J.N., Alsaka Y.A., Childers D.G. Variability in Closed Phase Analysis of Speech. ICA SSP 85, Tampa Fla, March 26-29, 1985, Vol 3 "New York, N.Y., 1985, 1089-1092.
7. Золотарев И.Д. Нестационарные процессы в резонансных усилителях фазово-импульсных измерительных систем. - Новосибирск: Наука. Сиб. отд., 1969. - С. 176.
8. Гитлин В.Б. К вопросу расчета формантных фильтров методом упрощенного преобразования Лапласа // Автоматические устройства учета и контроля. - Ижевск, 1977. - Вып. II. - С. 83-91.
9. Сметанин А.М. Исследование и разработка методов повышения точности измерений параметров формант и голосового источника. - Диссертация на соискание ученой степени к.т.н. Ижевск, 1980.
10. Lumis R.C. Speaker verification by computer Using Speech Intensity for Temporal Registration. IEEE Trans. Audio and Electroacoust. 1973, 21, №2, 80-89



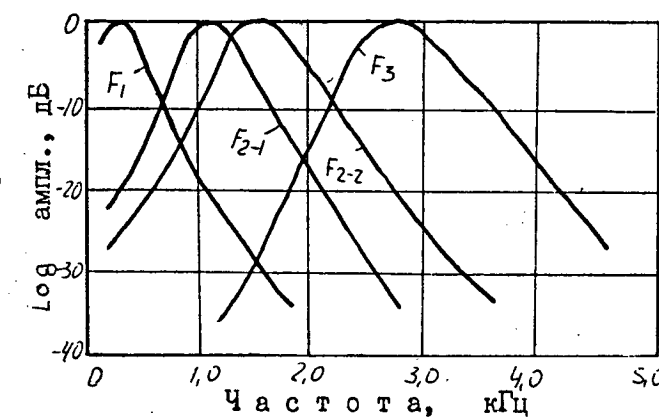
Фиг. 1



Фиг. 2



Фиг. 3



Фиг. 4