

Об одном алгоритме оценки формантных частот на интервале сомкнутых голосовых складок

Макаров И.С., кандидат технических наук

С теоретической точки зрения оценки формантных частот на интервале сомкнутых голосовых складок должны совпадать с резонансными частотами вокального тракта. Практически же формантный анализ на основе линейного предсказания, выполненный на таком интервале, зачастую приводит к нестабильным результатам: формантные частоты обнаруживают резкие нерегулярные скачки от периода к периоду, существенно искажающие ход формантной траектории, а кроме того, возможно появление так называемых «ложных» частотных оценок. В работе построен новый метод линейного предсказания, основанный на регуляризации траекторий коэффициентов предсказания. Сравнительный анализ данного алгоритма и стандартного ковариационного метода, проведённый на материале звукосочетаний «гласный + гласный» в произнесении мужчины и женщины, показал несомненное преимущество нового алгоритма с точки зрения стабильности, точности и физической правдоподобности получаемых формантных оценок.

Theoretically, formant frequencies estimated on the closed-glottis interval should coincide with the resonant frequencies of the vocal tract. Practically, LPC-based formant frequency analysis on the closed-glottis interval often results in abrupt, irregular formant trajectories. Appearance of so-called quasi-formant frequencies is also possible. In the article a new LPC method is constructed. This method is based on the regularization of linear prediction coefficients in time and results in much more regular and accurate formant frequencies estimates on the closed-glottis interval than the standard covariance LPC.

Введение

Необходимость автоматической оценки формантных частот вокального тракта по измеренному акустическому сигналу встаёт во многих задачах акустики речи:

- при автоматическом распознавании речи (Welling, Ney, 1998; Deng et al., 2006, 2007);
- при верификации и идентификации личности по голосу (Ramachandran et al., 1995);
- при кодировании речевого сигнала (Potamianos, Maragos, 1999);
- при синтезе речи (Pinto et al., 1989);
- при построении слуховых аппаратов (Mustafa, Bruce, 2006);
- при реабилитации пациентов с различными нарушениями речи и слуха (Pandey, Shah, 2009);
- при восстановлении формы речевого тракта по измеренному речевому сигналу (Леонов и др., 2005).

В мировой литературе решению задачи автоматической оценки формантных частот посвящено значительное число работ. Одним из наиболее популярных методов является метод оценки на основе модели линейного предсказания. В рамках данной модели отсчёт речевого сигнала в дискретный момент времени t определяется как линейная комбинация отсчётов того же сигнала в предыдущие моменты времени $t-1, t-2, \dots, t-p$ (Маркел, Грей, 1980):

$$s[t] = \sum_{i=1}^p a_i s[t-i] + e[t] \quad (1)$$

В соотношении (1): $\{a_i\}$ — коэффициенты линейного предсказания (называемые далее ЛП-коэффициентами), p — порядок модели, $e[t]$ — ошибка модели линейного предсказания (называемая также сигналом-остатком (residual signal) модели линейного предсказания). Для определения ЛП-коэффициентов по измеренному сегменту речи используется один из специальных алгоритмов, разработанных в рамках теории линейного предсказания. Обычно это автокорреляционный или ковариационный методы линейного предсказания (Маркел, Грей, 1980).

Как только ЛП-коэффициенты вычислены по анализируемому сегменту речи, формантные частоты, соответствующие данному сегменту, определяются либо по пикам ЛП-спектра, либо по полюсам ЛП-полинома (Маркел, Грей, 1980). Известны и более сложные алгоритмы: например, определение формантных частот по пикам групповой задержки ЛП-фильтра (Murthy, Yegnanarayana, 1991) или с помощью интегрирования по специальным контурам на комплексной плоскости (Snell, Milinazzo, 1993).

Хотя анализ методом линейного предсказания (и, как следствие, оценка формантных частот) может осуществляться асинхронно с основным тоном (ОТ), желательно уметь определять формантные частоты на каждом периоде ОТ, причём на временном интервале, соответствующем периоду сомкнутых голосовых складок. Поскольку на таком интервале, в основном, присутствуют свободные формантные колебания, ожидается, что соответствующие частотные оценки будут в наибольшей степени близки истинным формантным частотам речевого тракта (Krishnamurthy, Childers, 1986). Поскольку интервал сомкнутых голосовых складок зачастую оказывается очень коротким (длительность менее 3 мс), для проведения анализа используется ковариационный метод линейного предсказания в прямоугольном временном окне.

Метод оценки формантных частот на интервале сомкнутых голосовых складок сталкивается с двумя серьёзными проблемами. Во-первых, определение такого интервала непосредственно по акустическому сигналу является задачей нетривиальной. В мировой литературе известно множество методов, автоматически определяющих такой интервал (Parthasarathy, 1987; Fu, Murphy, 2006; Леонов и др., 2009), однако, судя по всему, задача окончательно не решена. Основная проблема связана с тем, что у некоторых людей интервал сомкнутых голосовых складок может быть слишком коротким для осуществления ЛП-анализа или даже отсутствовать совсем (последнее особенно характерно для женских и детских голосов). Чтобы обойти эту проблему, вводят представление об «эффективном» интервале сомкнутых голосовых складок (т.е. таком интервале, на котором голосовая щель фактически может быть открыта, однако колебания из трахеи, бронхов и лёгких в речевой тракт практически не проникают), причём определяют этот интервал с помощью искусственных приёмов предобработки речи (Akande, Murphy, 2005; Nordstrom et al., 2008) или с помощью специальных алгоритмов выбора этого интервала (Miyoshi et al., 1987; Ma et al., 1993; Magi et al., 2009). Другой путь, описанный в литературе, состоит в том, что сначала по сигналу оценивают моменты возбуждения вокального тракта голосовым источником, после чего определяют интервал сомкнутых голосовых складок как период времени длительностью 20–30% от периода ОТ, расположенный сразу же после момента возбуждения (Yegnanarayana, Veldhuis, 1998).

Вторая проблема оценки формантных частот на интервале сомкнутых голосовых складок связана с малой длительностью такого интервала (несколько миллисекунд). Хорошо известно (Stoica et al., 1989), что чем короче интервал ЛП-анализа, тем сильнее коэффициенты линейного предсказания оказываются зависимыми от условий порождения и регистрации речевого сигнала. Присутствие в сигнале посторонних колебаний даже очень малой амплитуды (будь то внешний шум, колебания из подсвязочных областей или влияние амплитудно-частотной характеристики микрофона) может привести к тому, что ЛП-коэффициенты будут определяться с очень большими погрешностями. Применительно к оценке формантных частот это означает, что частоты, вычисленные по таким коэффициентам, будут сильно отличаться от истинных значений резонансов речевого тракта. Практически это проявляется в скачкообразных нерегулярных изменениях оценок формантных частот от периода к периоду. При этом возможно появление «ложных» формантных оценок, не имеющих отношения к физической стороне процесса и обусловленных исключительно неустойчивостью процедуры ковариационного анализа к паразитным шумам. Возможна также ситуация пропуска «истинной» формантной оценки, что опять же связано с проблемой короткого интервала анализа.

В мировой литературе описано несколько попыток решения этой проблемы. В работе (Cranen, Voves, 1987) предложено перед осуществлением ковариационного ЛП-анализа усреднять по ансамблю сигналы на интервале закрытой голосовой щели, оценённые на соседних периодах ОТ. В работе (Yegnanarayana, Veldhuis, 1998) построен так называемый многоцикловый (multi-cycle) алгоритм, заключающийся в том, что сначала суммируются сигналы и соответствующие ковариационные матрицы, вычисленные внутри двух–трёх соседних периодов ОТ на интервалах сомкнутых голосовых складок, после чего ковариационный алгоритм применяется к результатам суммирования. В работе (Parthasarathy, 1987) ЛП-коэффициенты оцениваются не с помощью ковариационного метода линейного предсказания, а с помощью алгоритма, основанного на сингулярном разложении соответствующей ковариационной матрицы. Наконец, алгоритмы, построенные в (Miyoshi et al., 1987; Ma et al., 1993; Magi et al., 2009), основаны на взвешивании ковариационной матрицы перед осуществлением ЛП-анализа.

Мы протестировали все описанные алгоритмы, и ни один из них не показал явного преимущества в задаче оценки формантных частот на интервале сомкнутых голосовых складок; частично результаты тестирования описаны в (Леонов и др., 2009). Особенно ненадёжно все описанные алгоритмы работали в условиях наличия внешнего шума (белый шум с отношением сигнал-помеха > 15 дБ). Этот вывод согласуется с результатами работ

других авторов (Ramachandran et al., 1995; Sreenivas, Niederjohn, 1992). По этой причине возникла необходимость в создании более устойчивого алгоритма.

Алгоритм линейного предсказания с ограничениями на траектории ЛП-коэффициентов

Идея предлагаемого алгоритма заключается в следующем. Как указано во Введении, неустойчивость оценок ЛП-коэффициентов относительно паразитных колебаний на интервале сомкнутых голосовых складок проявляется в нерегулярных скачкообразных изменениях соответствующих оценок формантных частот от периода к периоду. Поскольку формантные частоты связаны с коэффициентами линейного предсказания взаимно-однозначной зависимостью (Kot et al., 1993), скачкообразное поведение частотных оценок предполагает скачкообразное изменение соответствующих ЛП-коэффициентов от периода к периоду. Возникает вопрос, нельзя ли вычислять коэффициенты линейного предсказания таким способом, чтобы по возможности не допустить скачков этих коэффициентов при переходе от периода к периоду. В этом случае можно надеяться на то, что траектории соответствующих формантных частот также не будут допускать резких, скачкообразных изменений.

Итак, идея заключается в построении такого алгоритма линейного предсказания, который не только определял бы ЛП-коэффициенты, но и ограничивал бы возможные изменения этих коэффициентов во времени, не допуская резких, скачкообразных изменений траекторий ЛП-коэффициентов.

Чтобы сформулировать задачу математически, перепишем соотношение (1) в векторно-матричной форме:

$$A\bar{a} = \bar{s} + \bar{e} \quad (2)$$

Здесь $\bar{a} = (a_1, a_2, \dots, a_p)^T$ — вектор ЛП-коэффициентов, T — значок транспонирования. Если длительность анализируемого сегмента речи в отсчётах равна K , то для ковариационного метода линейного предсказания $\bar{e} = (e[p+1], e[p+2], \dots, e[K])^T$ — вектор дискретных отсчётов сигнала-остатка, $\bar{s} = (s[p+1], s[p+2], \dots, s[K])^T$ — вектор дискретных отсчётов анализируемого сегмента речи, а матрица A определяется следующим образом (Ma et al., 1993):

$$A = \begin{pmatrix} s[p] & s[p-1] & \dots & s[1] \\ s[p+1] & s[p] & \dots & s[2] \\ \vdots & \vdots & \ddots & \vdots \\ s[K-1] & s[K-2] & \dots & s[K-p] \end{pmatrix} \quad (3)$$

Для всех случаев, практически важных для анализа речевого сигнала, система (2) будет переопределённой. По этой причине в классической теории линейного предсказания коэффициенты линейного предсказания определяют методом наименьших квадратов, т.е. как коэффициенты, минимизирующие ошибку модели линейного предсказания (2):

$$\tilde{a} = \arg \min_{\bar{a}} \|A\bar{a} - \bar{s}\|_2^2 \quad (4)$$

Здесь \tilde{a} — искомый вектор ЛП-коэффициентов, $\|\cdot\|_2^2$ — квадрат евклидовой нормы, определяемый для произвольного вектора размерности m как:

$$\|\bar{x}\|_2^2 = \sum_{i=1}^m x_i^2$$

Чтобы минимизировать скачки траекторий ЛП-коэффициентов, соотношение (4) можно модифицировать следующим образом:

$$\tilde{a}_n = \arg \min_{\tilde{a}_n} \left[\|A_n \bar{a}_n - \bar{s}_n\|_2^2 + \lambda \|\bar{a}_n - a_{n-1}\|_2^2 \right] \quad (5)$$

Соотношение (5) означает следующее: на n -ном периоде ОТ ищется такой вектор ЛП-коэффициентов \tilde{a}_n , который, с одной стороны, наилучшим образом аппроксимировал бы — в смысле модели (1) — сегмент речевого сигнала \bar{s}_n , измеренного на интервале сомкнутых голосовых складок n -ного периода (с матрицей A_n), и, с другой стороны, наименее отличался бы от вектора ЛП-коэффициентов \tilde{a}_{n-1} , уже определённого для предыдущего $(n-1)$ -го периода ОТ. Коэффициент λ в теории некорректных задач носит название «параметра регуляризации» (Тихонов, Арсенин, 1986). Он характеризует вклад второго члена (условие отсутствия резких скачков) в общий критерий (5), подлежащий минимизации.

Алгоритм определения коэффициентов линейного предсказания, основанный на (5), может быть легко реализован программно: на 1-м периоде ОТ определяется вектор ЛП-коэффициентов \tilde{a}_1 с помощью стандартного ковариационного метода путём минимизации соотношения (4). Подставляя \tilde{a}_1 в (5) и минимизируя это выражение по вектору \bar{a}_2 , получаем оценку \tilde{a}_2 коэффициентов линейного предсказания на втором периоде ОТ и т.д. Описанный алгоритм, однако, имеет один существенный изъян. Представим себе, что вектор \tilde{a}_1 на первом периоде ОТ определён с большой погрешностью. В этом случае все остальные ЛП-коэффициенты будут определяться из условия малого отличия от \tilde{a}_1 , т.е. также с большой погрешностью.

В дальнейшем будет построен алгоритм линейного предсказания, свободный от описанного изъяна. В рамках данного алгоритма будет определяться не один вектор ЛП-коэффициентов на некотором периоде ОТ, а сразу целая траектория коэффициентов линейного предсказания на протяжении нескольких периодов. Однако, прежде чем переходить к его построению, отметим, что соотношение (5) является одной из разновидностей так называемого регуляризованного линейного предсказания (Eckman et al., 2008).

Перейдём к математической формулировке алгоритма. Пусть $\bar{s}_1, \bar{s}_2, \dots, \bar{s}_N$ — векторы отсчётов речевого сигнала, измеренных на интервалах сомкнутых голосовых складок, принадлежащих последовательным периодам ОТ (1-й период, 2-й период, ..., N -ый период); $\bar{a}_1, \bar{a}_2, \dots, \bar{a}_N$ — векторы ЛП-коэффициентов, подлежащие оценке на соответствующих периодах; A_1, A_2, \dots, A_N — матрицы отсчётов сигнала на соответствующих периодах. Тогда задача, которую должен решать алгоритм, формулируется как нахождение последовательности векторов $\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_N$ из решения следующей оптимизационной задачи:

$$\{\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_N\} = \arg \min_{\{\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_N\}} \left[\sum_{n=1}^N \|A_n \bar{a}_n - \bar{s}_n\|_2^2 + \lambda \sum_{n=2}^N \|\bar{a}_n - \bar{a}_{n-1}\|_2^2 \right] \quad (6)$$

Соотношение (6) означает следующее: рассматриваются всевозможные последовательности векторов $\{\bar{a}_1, \bar{a}_2, \dots, \bar{a}_N\}$ (всевозможные траектории, описываемые вектором ЛП-коэффициентов). Из этих последовательностей выбирают такую, которая будет минимизировать критерий (6).

Очевидно, что прямой перебор таких траекторий невозможен. Форма критерия, подлежащего

минимизации, подсказывает нам, что поставленная задача является задачей динамического программирования (Беллман, 1960). Чтобы использовать идеологию динамического программирования, введём несколько вспомогательных функционалов:

$$W_i = W_i(\bar{a}_1, \bar{a}_2, \dots, \bar{a}_i) = \sum_{n=1}^i \|A_n \bar{a}_n - \bar{s}_n\|_2^2 + \lambda \sum_{n=2}^i \|\bar{a}_n - \bar{a}_{n-1}\|_2^2, i = 2, \dots, N$$

$$W_1 = \|A_1 \bar{a}_1 - \bar{s}_1\|_2^2,$$

$$D_i(\bar{a}_i, \bar{a}_{i-1}) = \|A_i \bar{a}_i - \bar{s}_i\|_2^2 + \lambda \|\bar{a}_i - \bar{a}_{i-1}\|_2^2, i = 2, \dots, N \quad (7)$$

Из (7) следует, что

$$\{\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_N\} = \arg \min_{\{\bar{a}_1, \bar{a}_2, \dots, \bar{a}_N\}} W_N(\bar{a}_1, \bar{a}_2, \dots, \bar{a}_N) \quad (8)$$

Кроме того, очевидно следующее равенство:

$$W_i(\bar{a}_1, \bar{a}_2, \dots, \bar{a}_i) = W_{i-1}(\bar{a}_1, \bar{a}_2, \dots, \bar{a}_{i-1}) + D_i(\bar{a}_i, \bar{a}_{i-1}), i = 2, \dots, N \quad (9)$$

Найдём минимум W_N . Используя (9), получаем следующий рекурсивный набор соотношений:

$$\begin{aligned} \min_{\{\bar{a}_1, \bar{a}_2, \dots, \bar{a}_N\}} W_N(\bar{a}_1, \bar{a}_2, \dots, \bar{a}_N) &= \min_{\{\bar{a}_1, \bar{a}_2, \dots, \bar{a}_N\}} [W_{N-1}(\bar{a}_1, \dots, \bar{a}_{N-1}) + D_N(\bar{a}_N, \bar{a}_{N-1})] = \\ &= \min_{\{\bar{a}_1, \bar{a}_2, \dots, \bar{a}_{N-1}\}} \left[W_{N-1} + \min_{\bar{a}_N} D_N(\bar{a}_N, \bar{a}_{N-1}) \right] = \\ &= \min_{\{\bar{a}_1, \bar{a}_2, \dots, \bar{a}_{N-1}\}} \left[W_{N-2} + D_{N-1}(\bar{a}_{N-1}, \bar{a}_{N-2}) + \min_{\bar{a}_N} D_N(\bar{a}_N, \bar{a}_{N-1}) \right] = \\ &= \min_{\{\bar{a}_1, \bar{a}_2, \dots, \bar{a}_{N-2}\}} \left[W_{N-2} + \min_{\bar{a}_{N-1}} \left(D_{N-1}(\bar{a}_{N-1}, \bar{a}_{N-2}) + \min_{\bar{a}_N} D_N(\bar{a}_N, \bar{a}_{N-1}) \right) \right] = \dots \\ &= \min_{\bar{a}_1} \left[W_1 + \min_{\bar{a}_2} \left(D_2(\bar{a}_2, \bar{a}_1) + \min_{\bar{a}_3} \left(D_3(\bar{a}_3, \bar{a}_2) + \dots \min_{\bar{a}_{N-1}} \right. \right. \right. \\ &\quad \left. \left. \left. \left(D_{N-1}(\bar{a}_{N-1}, \bar{a}_{N-2}) + \min_{\bar{a}_N} D_N(\bar{a}_N, \bar{a}_{N-1}) \right) \dots \right) \right) \right] \quad (10) \end{aligned}$$

Из (10) следует, что минимум функционала W_N может быть определён рекурсивно на основании следующих соотношений:

$$F_N(\bar{a}_{N-1}) = \min_{\bar{a}_N} D_N(\bar{a}_N, \bar{a}_{N-1})$$

$$F_{N-1}(\bar{a}_{N-2}) = \min_{\bar{a}_{N-1}} [D_{N-1}(\bar{a}_{N-1}, \bar{a}_{N-2}) + F_N(\bar{a}_{N-1})]$$

$$F_{N-2}(\bar{a}_{N-3}) = \min_{\bar{a}_{N-2}} [D_{N-2}(\bar{a}_{N-2}, \bar{a}_{N-3}) + F_{N-1}(\bar{a}_{N-2})]$$

...

$$F_1 = \min_{\bar{a}_1} [W_1 + F_2(\bar{a}_1)] \quad (11)$$

Введём ещё набор следующих функций:

$$\tilde{a}_N = f_N(\bar{a}_{N-1}) = \arg \min_{\bar{a}_N} D_N(\bar{a}_N, \bar{a}_{N-1})$$

$$\tilde{a}_{N-1} = f_{N-1}(\bar{a}_{N-2}) = \arg \min_{\bar{a}_{N-1}} [D_{N-1}(\bar{a}_{N-1}, \bar{a}_{N-2}) + F_N(\bar{a}_{N-1})]$$

$$\begin{aligned}\tilde{a}_{N-1} &= f_{N-1}(\bar{a}_{N-2}) = \arg \min_{\bar{a}_{N-1}} [D_{N-1}(\bar{a}_{N-1}, \bar{a}_{N-2}) + F_N(\bar{a}_{N-1})], \\ \tilde{a}_{N-2} &= f_{N-2}(\bar{a}_{N-3}) = \arg \min_{\bar{a}_{N-2}} [D_{N-2}(\bar{a}_{N-2}, \bar{a}_{N-3}) + F_{N-1}(\bar{a}_{N-2})], \\ &\dots \\ \tilde{a}_1 &= \arg \min_{\bar{a}_1} [W_1 + F_2(\bar{a}_1)]\end{aligned}\quad (12)$$

Теоретически, рекурсивные соотношения (11) и (12) позволяют найти решение оптимизационной задачи (6). Схема работает следующим образом. Начиная с периода ОТ с порядковым номером N , мы, используя соотношения (11), рекурсивно определяем выражение для $F_1(\bar{a}_1)$. Как только функция $F_1(\bar{a}_1)$ становится известной, мы определяем \tilde{a}_1 путем минимизации данной функции. Как только вектор ЛП-коэффициентов \tilde{a}_1 для первого периода ОТ определен, вектор \tilde{a}_2 для второго периода ОТ вычисляется, согласно (12), как $\tilde{a}_2 = f_2(\tilde{a}_1)$. Зная \tilde{a}_2 , мы определяем \tilde{a}_3 как $\tilde{a}_3 = f_3(\tilde{a}_2)$. Продолжая эту процедуру, мы последовательно определяем все искомые векторы ЛП-коэффициентов $\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_N$.

Для того чтобы практически использовать рекурсивные соотношения (11) и (12), необходимо найти аналитический вид функций F_N, F_{N-1}, \dots, F_1 . Определим сначала выражения для D_N и \tilde{a}_N . Поскольку соотношения (11) представляют собой задачи последовательной квадратичной минимизации, вектор \tilde{a}_N определяется из уравнения $\partial D_N(\bar{a}_N, \bar{a}_{N-1}) / \partial \bar{a}_N = 0$. Воспользуемся известным правилом дифференцирования евклидовой нормы по вектору (Лоусон, Хенсон, 1986):

$$\frac{d}{d\bar{x}} \|A\bar{x} - \bar{b}\|_2^2 = 2A^T(A\bar{x} - \bar{b}) \quad (13)$$

Отсюда получаем:

$$\frac{\partial D_N(\bar{a}_N, \bar{a}_{N-1})}{\partial \bar{a}_N} = 0 = (A_N^T A_N + \lambda E) \bar{a}_N - \lambda \bar{a}_{N-1} - A_N^T \bar{s}_N$$

В этом соотношении E — единичная матрица. Разрешая его относительно вектора ЛП-коэффициентов, получаем:

$$\tilde{a}_N = f_N(\bar{a}_{N-1}) = \arg \min_{\bar{a}_N} D_N(\bar{a}_N, \bar{a}_{N-1}) = (A_N^T A_N + \lambda E)^{-1} (\lambda \bar{a}_{N-1} + A_N^T \bar{s}_N) \quad (14)$$

Подставляя это выражение в соотношение для F_N (см. (11)), получаем:

$$\begin{aligned}F_N(\bar{a}_{N-1}) &= \left\| \lambda A_N (A_N^T A_N + \lambda E)^{-1} \bar{a}_{N-1} + (A_N (A_N^T A_N + \lambda E)^{-1} A_N^T - E) \bar{s}_N \right\|_2^2 + \\ &+ \lambda \left\| (\lambda (A_N^T A_N + \lambda E)^{-1} - E) \bar{a}_{N-1} + (A_N^T A_N + \lambda E)^{-1} A_N^T \bar{s}_N \right\|_2^2\end{aligned}\quad (15)$$

Таким образом получим аналитические выражения для F_{N-1}, \dots, F_1 и f_{N-1}, \dots, f_1 . Из (11) и (12) следует, что общая запись для \tilde{a}_i и F_i определяется следующими соотношениями:

$$\begin{aligned}\tilde{a}_i &= B_i \bar{a}_{i-1} + C_i \\ F_i(\bar{a}_{i-1}) &= \sum_{k=1}^{N-i+1} \left[\|P_k \bar{a}_{i-1} + Q_k\|_2^2 + \lambda \|R_k \bar{a}_{i-1} + S_k\|_2^2 \right] i = N, N-1, \dots, 2\end{aligned}\quad (16)$$

Здесь $B_i, C_i, P_{ik}, Q_{ik}, R_{ik}, S_{ik}$ — некоторые матрицы. Для $i = N$ эти матрицы определяются из (14) и (15):

$$\begin{aligned} B_N &= \lambda(A_N^T A_N + \lambda E)^{-1}, \\ C_N &= (A_N^T A_N + \lambda E)^{-1} A_N^T \bar{s}_N, \\ P_{N,1} &= \lambda A_N (A_N^T A_N + \lambda E)^{-1}, \\ Q_{N,1} &= \left[A_N (A_N^T A_N + \lambda E)^{-1} A_N^T - E \right] \bar{s}_N, \\ R_{N,1} &= \lambda (A_N^T A_N + \lambda E)^{-1} - E, \\ S_{N,1} &= (A_N^T A_N + \lambda E)^{-1} A_N^T \bar{s}_N \end{aligned} \quad (17)$$

Допустим, эти матрицы уже определены на i -том шаге рекурсии (т.е. известны для \tilde{a}_i и F_i). Определим $B_{i-1}, C_{i-1}, P_{i-1,k}, Q_{i-1,k}, R_{i-1,k}, S_{i-1,k}$. Пусть $I > 2$. Тогда, используя (13), получаем:

$$\begin{aligned} \frac{\partial}{\partial \bar{a}_{i-1}} \left(\|A_{i-1} \bar{a}_{i-1} - \bar{s}_{i-1}\|_2^2 + \lambda \|\bar{a}_{i-1} - \bar{a}_{i-2}\|_2^2 + F_i(\bar{a}_{i-1}) \right) &= 0 = \\ = A_{i-1}^T (A_{i-1} \bar{a}_{i-1} - \bar{s}_{i-1}) + \lambda (\bar{a}_{i-1} - \bar{a}_{i-2}) + \sum_{k=1}^{N-i+1} \left[P_k^T (P_k \bar{a}_{i-1} + Q_k) + \lambda R_k^T (R_k \bar{a}_{i-1} + S_k) \right] \end{aligned} \quad (18)$$

Разрешая это уравнение относительно \bar{a}_{i-1} и сравнивая полученное соотношение с первым уравнением (16), получаем:

$$\begin{aligned} B_{i-1} &= \lambda \left(A_{i-1}^T A_{i-1} + \sum_{k=1}^{N-i+1} [P_k^T P_k + \lambda R_k^T R_k] + \lambda E \right)^{-1} \\ C_{i-1} &= \left(A_{i-1}^T A_{i-1} + \sum_{k=1}^{N-i+1} [P_k^T P_k + \lambda R_k^T R_k] + \lambda E \right)^{-1} \\ &\quad \left(A_{i-1}^T \bar{s}_{i-1} - \sum_{k=1}^{N-i+1} [P_k^T Q_k + \lambda R_k^T S_k] \right) \end{aligned} \quad (19)$$

Подставляя эти соотношения во второе уравнение (16), получаем:

$$\begin{aligned} P_{i-1,k} &= P_k B_{i-1}, Q_{i-1,k} = P_k C_{i-1} + Q_k, R_{i-1,k} = R_k B_{i-1}, S_{i-1,k} = R_k C_{i-1} + S_k, k = \\ &= 1, \dots, N-i+1 \\ P_{i-1, N-i+2} &= A_{i-1} B_{i-1}, Q_{i-1, N-i+2} = A_{i-1} C_{i-1} - \bar{s}_{i-1}, R_{i-1, N-i+2} = B_{i-1} - E, S_{i-1, N-i+2} = \\ &= C_{i-1} \end{aligned} \quad (20)$$

Аналогичным образом можно определить соответствующие матрицы для случая

$I = 2$. Нетрудно получить и выражение для \tilde{a}_1 :

$$\begin{aligned} \tilde{a}_1 &= \left[A_1^T A_1 + \sum_{k=1}^{N-1} (P_{2,k}^T P_{2,k} + \lambda R_{2,k}^T R_{2,k}) \right]^{-1} \\ &\quad \left[A_1^T \bar{s}_1 - \sum_{k=1}^{N-1} (P_{2,k}^T Q_{2,k} + \lambda R_{2,k}^T S_{2,k}) \right] \end{aligned} \quad (21)$$

Полученные формулы позволяют построить алгоритм, решающий исходную задачу (6). Алгоритм работает следующим образом.

1. Сначала по акустическому сигналу определяется соответствующий контур периодов основного тона. Для решения этой задачи мы используем алгоритм, построенный в (Цыплихин, 2007).
2. Для дальнейшего анализа выделяется последовательность из N периодов ОТ, при этом для каждого периода определяются интервалы сомкнутых голосовых складок. Для решения этой задачи мы используем алгоритм из (Yegnanarayana, Veldhuis, 1998).
3. На каждом из интервалов сомкнутых голосовых складок мы измеряем соответствующие акустические сигналы $\bar{s}_1, \dots, \bar{s}_N$ и матрицы A_1, \dots, A_N .
4. С помощью (17) определяем матрицы $B_N, C_N, P_{N,1}, Q_{N,1}, R_{N,1}, S_{N,1}$ для N -го периода ОТ.
5. С помощью соотношений (19) и (20) рекурсивно определяем выражения для $B_p, C_p, P_{ik}, Q_{ik}, R_{ik}, S_{ik}$ ($i = N, N-1, \dots, 2$).
6. Зная $P_{2,k}, Q_{2,k}, R_{2,k}, S_{2,k}$ определяем \tilde{a}_1 с помощью (21).
7. Зная \tilde{a}_1 , последовательно определяем $\tilde{a}_2, \tilde{a}_3, \dots, \tilde{a}_N$ с помощью первого соотношения (16). Полученная последовательность и будет минимизировать функционал (6).

Эксперименты

В качестве тестового материала мы использовали звукосочетания /IA/, /AI/, /IU/, /UI/, /AU/, /UA/ в произнесении двоих дикторов — носителей американского английского языка: мужчины (среднее значение частоты основного тона = 120 Гц) и женщины (среднее значение частоты основного тона = 190 Гц). Соответствующие звуковые файлы были взяты из базы данных X-ray Microbeam Speech Production Database, task 15 для дикторов jw12 (мужчина) и jw13 (женщина). Исходные акустические файлы были записаны с частотой дискретизации 21379 Гц при отношении сигнал-шум примерно 30 дБ (Westbury, 1994). Эти звуковые файлы анализировались построенным алгоритмом, и полученные формантные треки отображались на сонограммах соответствующих сигналов для визуальной оценки того, насколько точно и надёжно алгоритм отслеживает соответствующие формантные частоты. Для сравнения результатов мы использовали стандартный ковариационный метод линейного предсказания на интервале сомкнутых голосовых складок применительно ко всем акустическим файлам, описанным выше.

Анализ осуществлялся следующим образом. Сначала все речевые сигналы ресэмплировались на частоту дискретизации 8 кГц с 16-битным квантованием амплитуды каждого временного отсчёта. Затем по этим сигналам определялись моменты возбуждения вокального тракта голосовым источником с помощью метода, построенного в (Цыплихин, 2007). Временной интервал между двумя соседними голосовыми возбуждениями принимался нами за текущий период основного тона. Длительность интервала сомкнутых голосовых складок принималась равной двум миллисекундам, при этом начало данного интервала совмещалось с первым отсчётом, следующим за моментом возбуждения (данная стратегия выбора интервала сомкнутых складок была описана в (Yegnanarayana, Veldhuis, 1998)). Речевой сигнал не предсказался.

Построенный алгоритм линейного предсказания, а также стандартный ковариационный метод были применены для оценки формантных частот на выделенных интервалах сомкнутых голосовых складок. В обоих методах в качестве временного окна анализа использовалось прямоугольное окно. Порядок модели для обоих алгоритмов был равен 9. Параметр регуляризации λ во всех случаях полагался равным 0.05. После определения ЛП-векторов формантные частоты и ширины определялись по полюсам соответствующих ЛП-полиномов. При этом из анализа удалялись полюсы с частотами, меньшими

200 Гц, полюсы с ширинами, большими 600 Гц, а также полюс на частоте Найквиста (частота, равная половине частоты дискретизации).

На рис. 1 представлена сонограмма звуко сочетания /AI/ в произнесении диктора-мужчины вместе с формантными треками, определёнными с помощью ковариационного метода линейного предсказания.

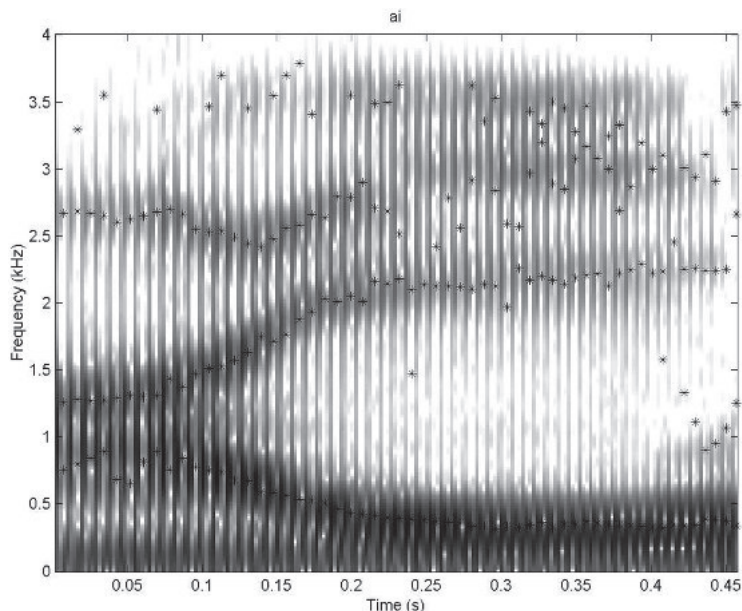


Рис. 1. Сонограмма /AI/ вместе с формантными треками; ковариационный метод; диктор-мужчина

Как видно из рисунка, оценки формантных частот обнаруживают нерегулярные резкие скачки от периода к периоду. Для первой форманты такие скачки наблюдаются в начале звукового файла (временной интервал в окрестности метки 0.05 сек), на протяжении оставшейся части звуко сочетания траектория $F1$ сравнительно гладкая. Резкие скачки оценок второй форманты практически не наблюдаются, за исключением отдельных периодов (два периода в окрестности метки 0.3 сек). Трек оценок третьей форманты непрерывен примерно до временной метки 0.2 сек (гласный /A/ и формантный переход на /I/). Затем траектория $F3$ фактически распадается на последовательность нерегулярных скачков. Отметим, что на основной части временного интервала, соответствующего гласному /I/, оценки $F3$ существенно отличаются от «истинного» хода третьей форманты. Наконец, трек четвёртой форманты фактически не просматривается.

Помимо скачкообразного поведения оценок формантных частот, на рис. 1 между треками $F1$ и $F2$ наблюдаются «ложные» форманты – в окрестности метки 0.24 сек и в конце звукового файла (0.4–0.45 сек).

На рис. 2 показана сонограмма того же звуко сочетания в произнесении диктора-женщины.

Макаров И.С.

Об одном алгоритме оценки формантных частот на интервале сомкнутых голосовых складок

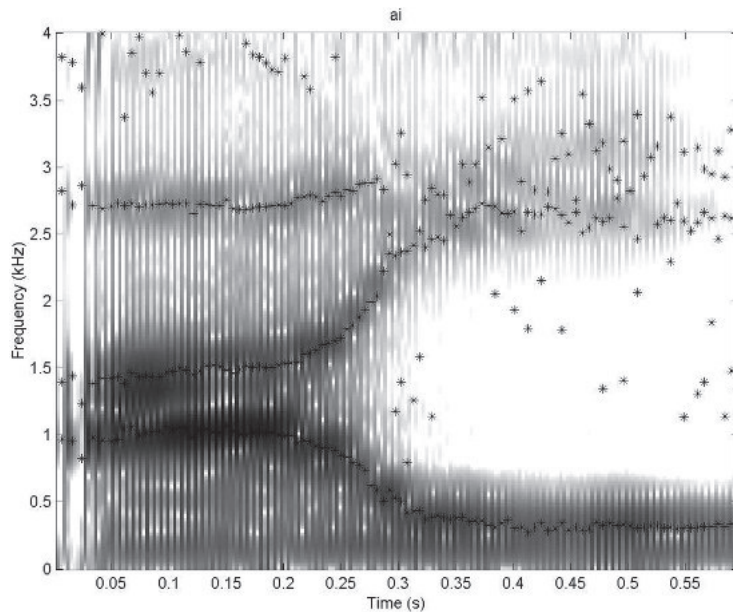


Рис. 2. Сонограмма /AI/ вместе с формантными треками; ковариационный метод; диктор-женщина

Очевидны резкие нерегулярные скачки, существенно искажающие ход формантных траекторий (особенно для второй и третьей форманты). Начиная с метки 0.28 сек, F3-трек практически не просматривается. Помимо этого, на интервале 0.25–0.6 сек между траекториями второй и третьей формант наблюдается большое количество ложных формант.

На рис. 3 и 4 представлены формантные треки звукосочетания /AI/, оценённые новым алгоритмом линейного предсказания для, соответственно, мужского и женского голосов.

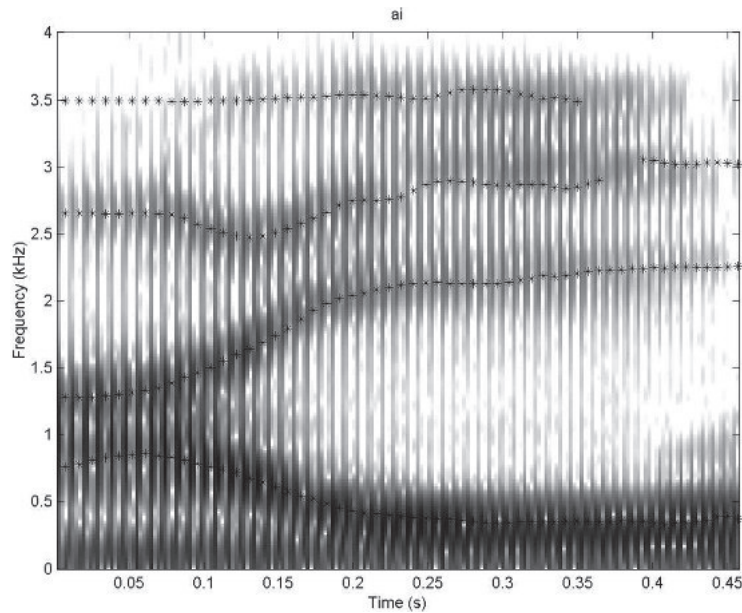


Рис. 3. Сонограмма /AI/ вместе с формантными треками; новый ЛП-алгоритм; диктор-мужчина

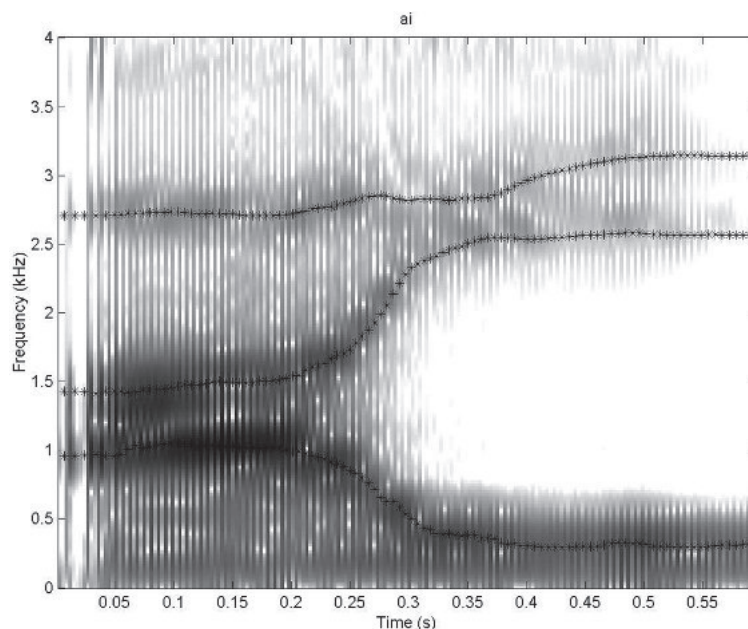


Рис. 4. Сонограмма /AI/ вместе с формантными треками; новый ЛП-алгоритм; диктор-женщина

Сравнивая рис. 1 и 3 с рис. 2 и 4 соответственно, мы видим, что новый алгоритм линейного предсказания обеспечивает гораздо более гладкие и физически правдоподобные траектории оценок формантных частот, чем стандартный ковариационный метод. В частности, не наблюдается ни одной ложной формантной оценки, а треки второй и третьей формант определяются с высокой точностью. Это особенно очевидно для женского голоса на временном интервале 0.35–0.6 сек (траектории второй и третьей формант). Для мужского голоса на интервале 0.37–0.4 сек для третьей форманты наблюдается пропуск нескольких оценок. Этот пропуск, связанный с тем, что ширины соответствующих полюсов на данном временном интервале оказались больше 600 Гц, может быть легко исправлен в автоматическом алгоритме построения формантных треков. Интересно отметить, что для мужского голоса траектория четвёртой форманты определяется весьма надёжно до метки 0.35 сек.

На рис. 5 и 6 показаны сонограммы и формантные треки для звукосочетания /AU/ в произнесении мужчины и женщины.

Для мужского голоса ковариационный анализ сравнительно успешно оценил треки всех трёх формант. Напротив, для женского голоса оценки второй форманты (особенно на интервалах 0–0.2 сек и 0.3–0.65 сек) демонстрирует нерегулярное скачкообразное поведение, существенно затемняющее «истинный» ход F_2 -траектории. Наконец, трек третьей форманты — начиная с метки 0.25 сек — практически не просматривается.

Макаров И.С.

Об одном алгоритме оценки формантных частот на интервале сомкнутых голосовых складок

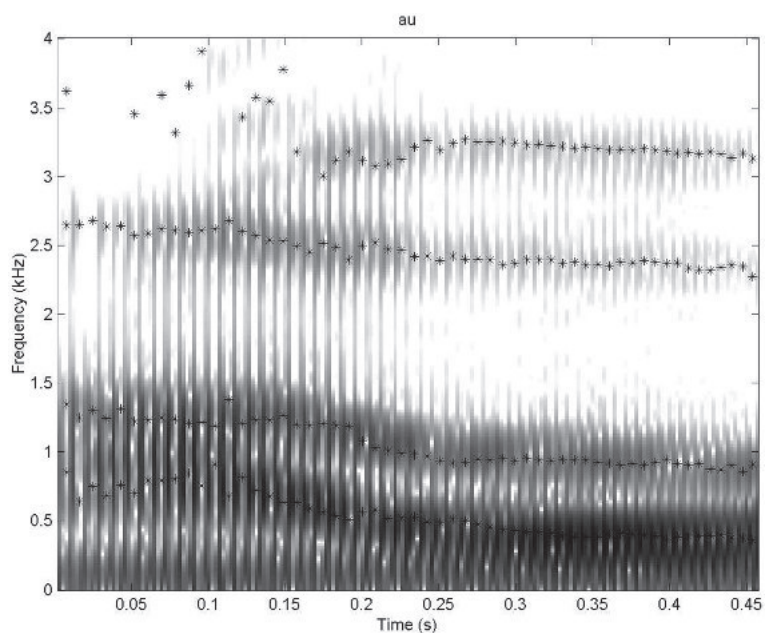


Рис. 5. Сонограмма /AU/ вместе с формантными треками; ковариационный метод; диктор-мужчина

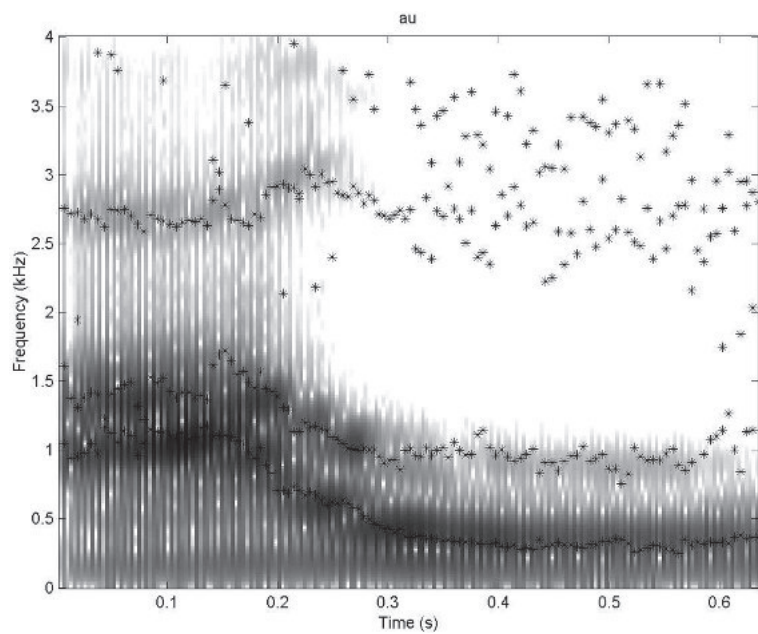


Рис. 6. Сонограмма /AU/ вместе с формантными треками; ковариационный метод; диктор-женщина

Соответствующие результаты, полученные с помощью нового алгоритма, показаны на рис. 7 и 8.

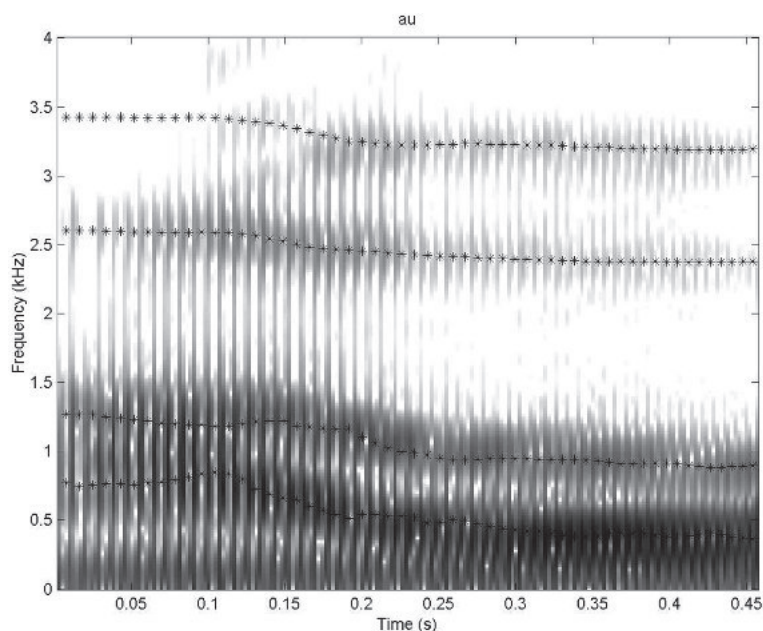


Рис. 7. Сонограмма /AU/ вместе с формантными треками; новый ЛП-алгоритм; диктор-мужчина

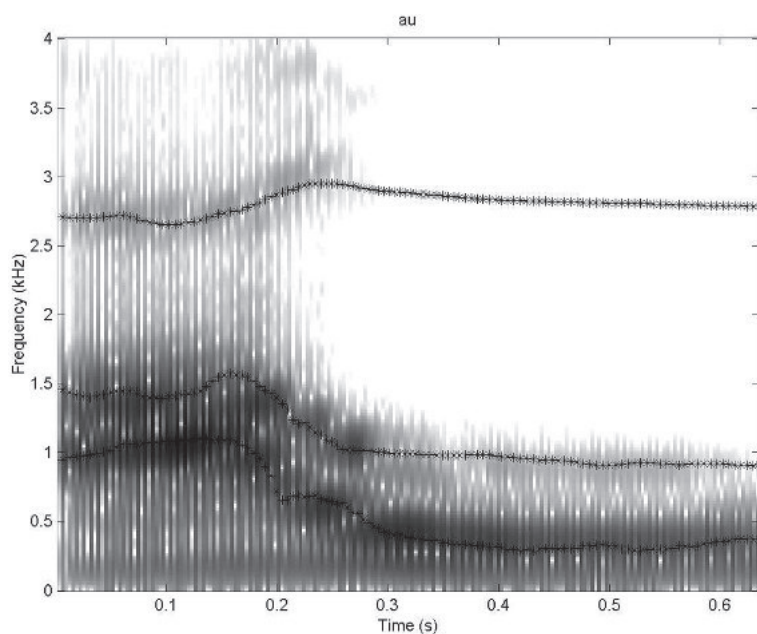


Рис. 8. Сонограмма /AU/ вместе с формантными треками; новый ЛП-алгоритм; диктор-женщина

Как и в случае звукосочетания /AI/, новый алгоритм надёжно оценил траектории трёх формантных частот как для мужского, так и для женского голоса. Например, поведение третьей форманты на интервале 0.25–0.65 сек для женского голоса было отслежено очень чётко (рис. 8), в то время как стан-



Макаров И.С.

Об одном алгоритме оценки формантных частот на интервале сомкнутых голосовых складок

дартный ковариационный метод не смог на том же интервале построить формантной траектории.

На рис. 9 и 10 показаны сонограммы и формантные треки для звукосочетания /IU/.

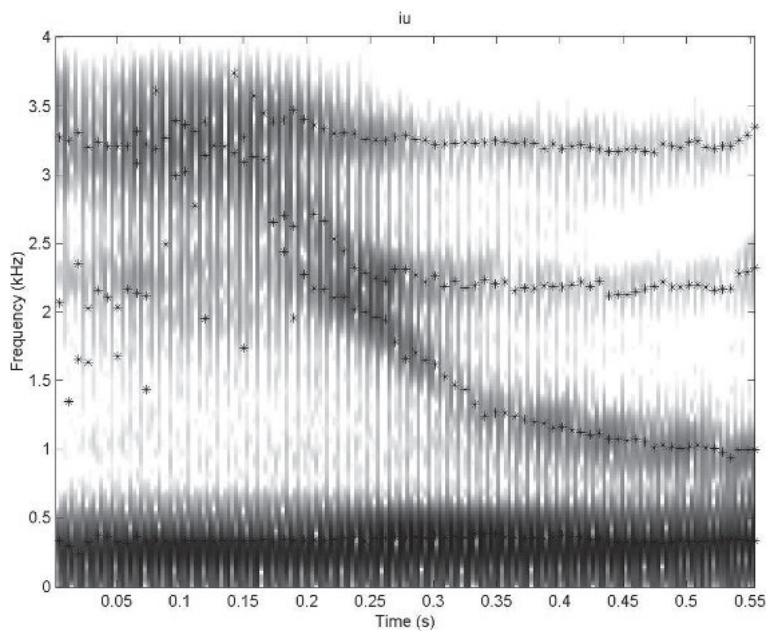


Рис. 9. Сонограмма /IU/ вместе с формантными треками; ковариационный метод; диктор-мужчина

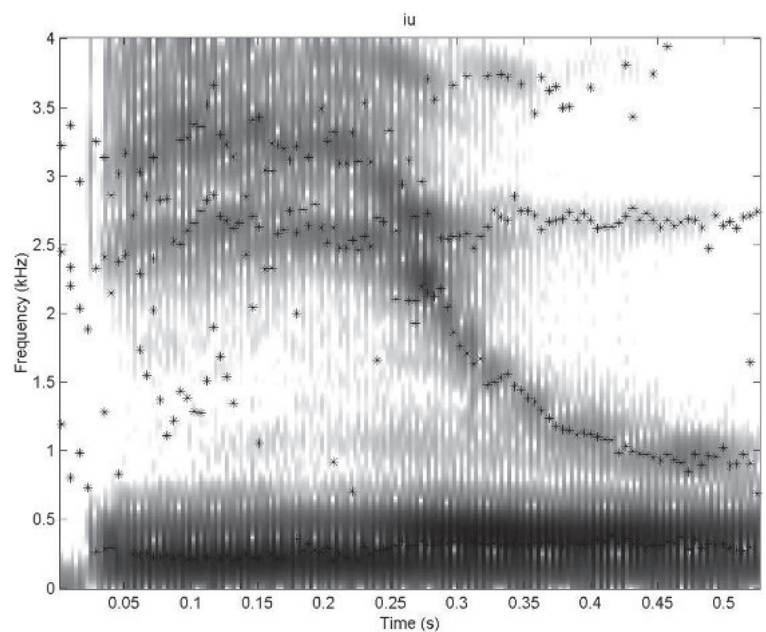


Рис. 10. Сонограмма /IU/ вместе с формантными треками; ковариационный метод; диктор-женщина

Видно, что как для мужского, так и для женского голосов треки второй и третьей формант, вычисленные с помощью стандартного ковариационного метода, носят крайне нерегу-



Макаров И.С.

Об одном алгоритме оценки формантных частот на интервале сомкнутых голосовых складок

лярный скачкообразный характер, а на интервале до 0.25 сек фактически вообще не просматриваются. Напротив, новый ЛП-алгоритм определяет соответствующие траектории очень чётко (рис. 11 и 12).

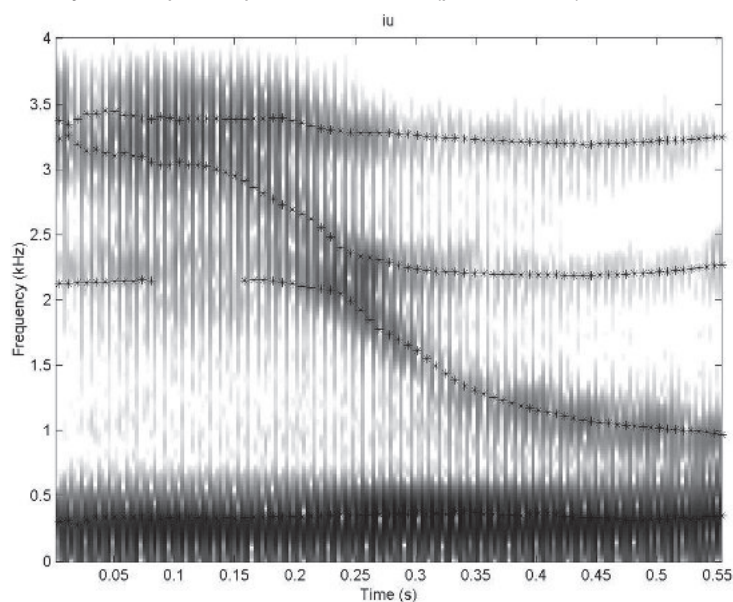


Рис. 11. Сонограмма /IU/ вместе с формантными треками; новый ЛП-алгоритм; диктор-мужчина

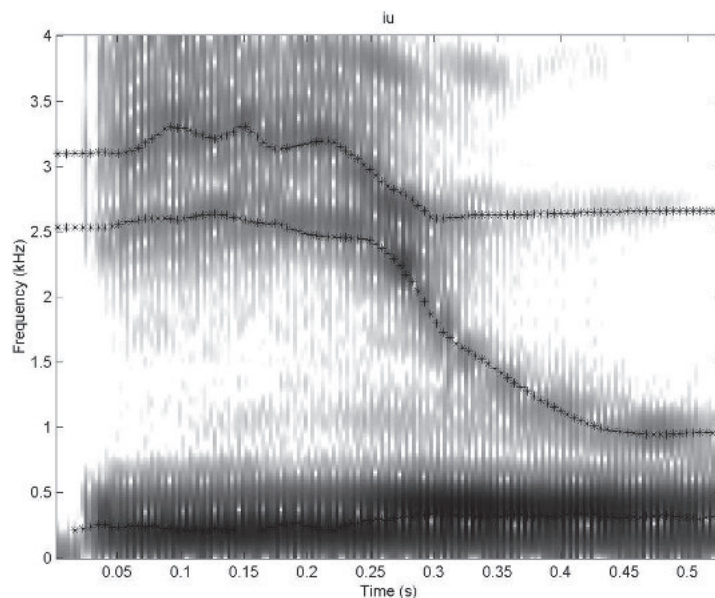


Рис. 12. Сонограмма /IU/ вместе с формантными треками; новый ЛП-алгоритм; диктор-женщина

Как видно из рис. 11, на интервале 0.1–0.15 сек новый алгоритм пропустил несколько оценок второй форманты. Как и в случае звукосочетания /AI/, этот пропуск связан с тем, что ширина соответствующего полюса на этом интервале превысила 600 Гц (таким образом, фактически оценки частоты второй



Макаров И.С.

Об одном алгоритме оценки формантных частот на интервале сомкнутых голосовых складок

форманты там присутствуют). Для устранения этого недостатка можно использовать разные стратегии, наиболее простая из которых — увеличить порог для допустимой ширины форманты. В любом случае, поскольку такой пропуск включает в себя лишь несколько частотных оценок, его легко отследить и компенсировать в автоматическом формантном трекаре.

Обсуждение

Статья посвящена построению нового алгоритма линейного предсказания с ограничениями на допустимые траектории ЛП-коэффициентов применительно к задаче оценивания формантных частот на очень коротких интервалах (соответствующих интервалам сомкнутых голосовых складок). Сравнительный анализ данного алгоритма и стандартного ковариационного метода, проведённый на звукосочетаниях «гласный + гласный» в произнесении мужчины и женщины, показал, что во всех случаях формантные треки, построенные с помощью нового алгоритма, оказываются гораздо более гладкими, точными и физически правдоподобными, чем траектории, определённые стандартным методом. В частности, не наблюдается ни нерегулярных скачкообразных изменений формантных частот от периода к периоду, столь характерных для линейного предсказания на интервале сомкнутых голосовых складок, ни появления «ложных» формантных оценок.

Построенный алгоритм существенно зависит от параметра регуляризации λ . Чем больше значение этого параметра, тем существеннее вклад члена, определяющего непрерывность ЛП-траектории, в общий критерий. В наших экспериментах мы использовали постоянное значение этого параметра, равное 0.05. Возникает вопрос, как отразится на результатах формантного анализа вариация значений этого параметра.

Выше был отмечен пропуск нескольких оценок второй форманты для звукосочетания /IU/ в произнесении диктора-мужчины (рис. 11). Если положить параметр регуляризации равным 0.09, получим следующий результат (рис. 13):

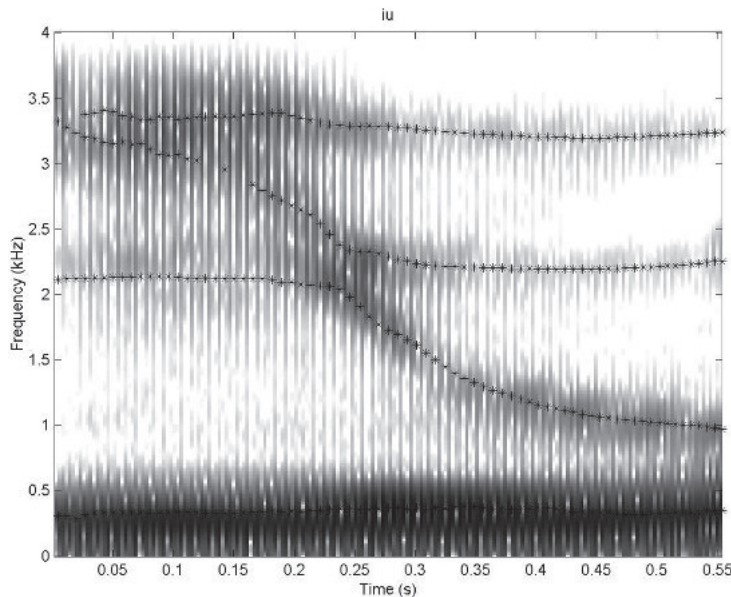


Рис. 13. Сонограмма /IU/ вместе с формантными треками; новый ЛП-алгоритм; диктор-мужчина

Сравнивая рис. 11 и 13, мы видим, что пропуск формантных оценок на F_2 -траектории исчез. Вместе с тем это привело к некоторому смещению трека формантных оценок по частоте (по сравнению с «истинной» формантной траекторией), а также к возникновению трёх пропусков на F_3 -треке.



Рассмотрим другой пример. На рис. 14 представлена сонограмма звукосочетания /IA/ для женского голоса, полученная ковариационным методом, а на рис. 15 — новым ЛП-алгоритмом с параметром регуляризации = 0.05.

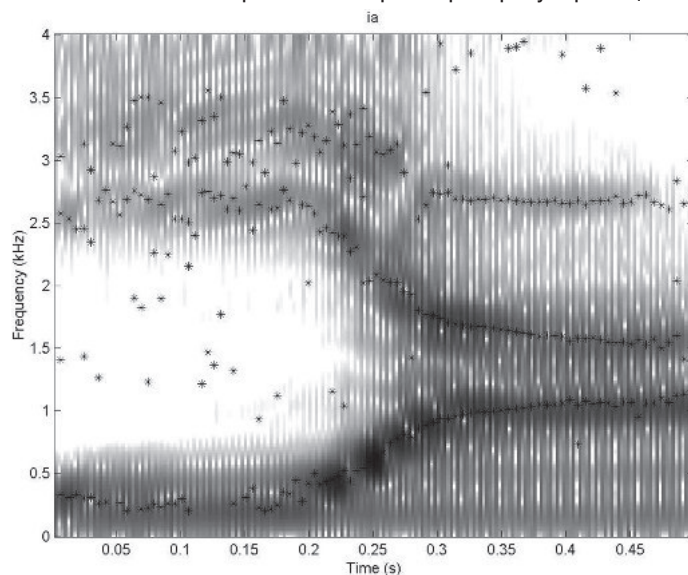


Рис. 14. Сонограмма /IA/ вместе с формантными треками; ковариационный метод; диктор-женщина

Из сравнения рис. 13 и 14 видно, что новый алгоритм определил треки второй и третьей формант гораздо лучше, чем стандартный ковариационный анализ. Однако на интервале 0.22–0.27 сек на траектории третьей форманты наблюдаются скачки, искажающие ход трека. Чтобы избавиться от этого скачка, мы увеличили значение параметра регуляризации с 0.05 до 0.8. Результирующие формантные треки показаны на рис. 15.

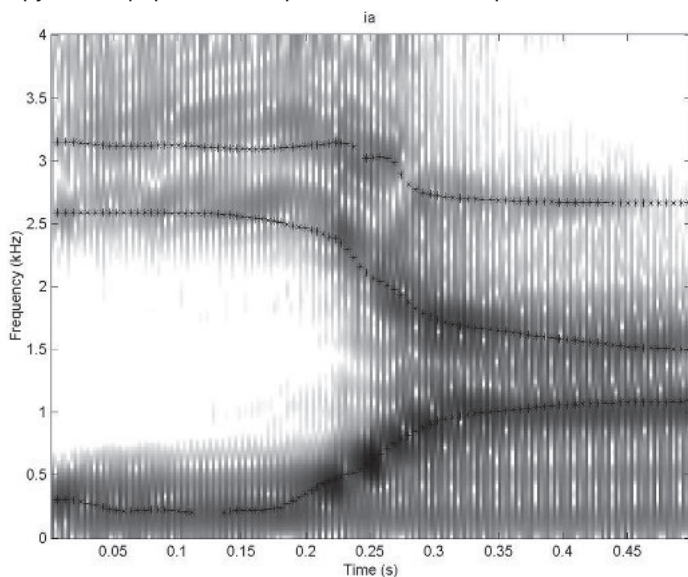


Рис. 15. Сонограмма /IA/ вместе с формантными треками; новый ЛП-алгоритм; диктор-женщина



Видно, что скачки практически полностью исчезли. Вместе с тем, усиление веса критерия непрерывности привело к некоторому смещению формантных оценок по частоте (особенно очевидному для второй и третьей форманты на интервале до 0.15 сек). Это плата за непрерывность формантных траекторий.

Из приведённых примеров следует, что, варьируя параметр регуляризации, можно добиться большей или меньшей степени непрерывности и гладкости формантных траекторий за счёт некоторого смещения мгновенных формантных оценок. Возникает вопрос, можно ли создать алгоритм, который бы автоматически подбирал значения параметра регуляризации, достигающие оптимального компромисса между требованием непрерывности формантных траекторий и малой ошибкой аппроксимации речевого сигнала моделью линейного предсказания (1).

В теории некорректных задач одним из методов, решающих данную проблему, является метод подбора параметра регуляризации по невязке (Тихонов, Арсенин, 1986). Пусть δ – погрешность измерения речевого сигнала, $\tilde{a}_{k,\lambda}$ – k -тый ЛП-вектор из оптимальной последовательности векторов, определённых с помощью построенного алгоритма при заданном значении параметра регуляризации λ . Тогда сущность метода заключается в выборе такого значения λ^* , которое бы удовлетворяло следующему соотношению:

$$\lambda^* = \arg \min_{\lambda} \left[\frac{1}{N} \sum_{k=1}^N \|A_k \tilde{a}_{k,\lambda} - \bar{s}_k\|_2 - \delta \right]$$

Практически решение этого нелинейного уравнения можно получить следующим образом. Для разных заранее заданных значений параметра регуляризации λ мы находим с помощью нашего алгоритма последовательности ЛП-векторов. Из этих последовательностей мы выбираем такую, которая минимизирует данное соотношение. Значение параметра регуляризации, соответствующее выбранной оптимальной последовательности, и будет искомым значением λ^* . В настоящее время алгоритм, реализующий автоматический выбор оптимального значения параметра регуляризации, находится в процессе доработки и тестирования.

Построенный алгоритм определяет последовательность ЛП-коэффициентов из условия минимума функционала (6). Можно модифицировать данный функционал, записав его в следующей форме:

$$\{\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_N\} = \arg \min_{\{\bar{a}_1, \bar{a}_2, \dots, \bar{a}_N\}} \left[\sum_{n=1}^N \|A_n \bar{a}_n - \bar{s}_n\|_2^2 + \lambda \sum_{n=2}^N \|\bar{a}_n - 2\bar{a}_{n-1} + \bar{a}_{n-2}\|_2^2 \right]$$

Отличие данного функционала от функционала (6) заключается в том, что здесь мы требуем не непрерывности траекторий ЛП-коэффициентов, а их гладкости. Используя метод, построенный в данной работе, можно модифицировать его применительно к минимизации данного функционала. Пока неясно, будет ли решение этой минимизационной задачи иметь какое-то преимущество перед алгоритмом, построенным нами выше. Этот вопрос сейчас находится на стадии изучения.

Алгоритм линейного предсказания, построенный в этой работе, показал хорошие результаты на материале звукосочетаний, произнесённых мужчиной и женщиной при отношении сигнал-шум около 30 дБ. Необходимо исследование того, насколько точно и надёжно формантные частоты определяются в условиях шумов. В настоящее время эта проблема нами исследуется.

Заключение

Работа посвящена построению нового алгоритма линейного предсказания с ограничениями на траектории ЛП-коэффициентов. В основе алгоритма лежит минимизация некоторого критерия, включающего в себя ошибку аппроксимации речевого сигнала моделью



линейного предсказания и невязку между ЛП-векторами, вычисленными на соседних периодах основного тона. Минимизационная задача решается с помощью рекурсивной матричной схемы, построенной в рамках идеологии динамического программирования.

Алгоритм протестирован на звукосочетаниях «гласный + гласный» для мужского и женского голосов. Траектории оценок формантных частот, получаемых с помощью данного алгоритма, оказываются значительно лучше (по таким параметрам, как отсутствие резких скачков и ложных формант, физическая адекватность решения), чем траектории, определённые с помощью ковариационного метода линейного предсказания.

Литература

1. *Welling L., Ney H.* Formant Estimation for Speech Recognition. // IEEE Trans. Speech, and Audio Process. 1998. Vol. 6. № 1. P. 36–48.
2. *Deng L., A. Acero, and Bazzi I.* Tracking Vocal Tract Resonances Using a Quantized Nonlinear Function Embedded in a Temporal Constraint. // IEEE Trans. Audio, Speech, and Language Process. 2006. Vol. 14. № 2. P. 425–434.
3. *Deng L., Lee L., Attias H. and Acero A.* Adaptive Kalman Filtering and Smoothing for Tracking Vocal Tract Resonances Using a Continuous-Values Hidden Dynamic Model. // IEEE Trans. Audio, Speech, and Language Process. 2007. Vol. 15. № 1. P. 13–23.
4. *Ramachandran R., Zilovich M. and Mammone R.* A Comparative Study of Robust Linear Predictive Analysis Methods with Application to Speaker Identification. // IEEE Trans. Speech, and Audio Process. 1995. Vol. 3. № 2. P. 117–125.
5. *Potamianos A., Maragos P.* Speech Analysis and Synthesis Using an AM-FM Modulation Model. // Speech Communication. 1999. Vol. 28. P. 195–209.
6. *Pinto N., Childers D. and Lalwani A.* Formant Speech Synthesis: Improving Production Quality. // IEEE Trans. Acoust., Speech, and Signal Process. 1989. Vol. 37. № 12. P. 1870–1887.
7. *Mustafa K., Bruce I.* Robust Formant Tracking for Continuous Speech with Speaker Variability. // IEEE Trans. Audio, Speech, and Language Process. 2006. Vol. 14. № 2. P. 435–444.
8. *Pandey P., Shah M.* Estimation of Places of Articulation During Stop Closures of Vowel-Consonant-Vowel Utterances. // IEEE Trans. Audio, Speech, and Language Process. 2009. Vol. 17. № 2. P. 277–285.
9. *Леонов А.С., Макаров И.С., Сорокин В.Н., Цыплихин А.И.* Кодовая книга для речевых обратных задач. // Информационные процессы. 2005. Т. 5. № 2. С. 101–119.
10. *Маркел Дж., Грей А.* *Линейное предсказание речи.* М.: Связь, 1980.
11. *Murthy H., Yegnanarayana B.* Formant Extraction from Group Delay Function. // Speech Communication. 1991. Vol. 10. P. 209–221.
12. *Snell R., Milinazzo F.* Formant Location from LPC Analysis Data. // IEEE Trans. Speech, and Audio Process. 1992. Vol. 1. № 2. P. 129–134.
13. *Krishnamurthy A., Childers D.* Two-Channel Speech Analysis. // IEEE Trans. Acoust., Speech, and Signal Process. 1986. Vol. 34. № 4. P. 730–743.
14. *Parthasarathy S.* Excitation-Synchronous Modeling of Voiced Speech. // IEEE Trans. Acoust., Speech, and Signal Process. 1987. Vol. 35. № 9. P. 1241–1249.
15. *Fu Q., Murphy P.* Robust Glottal Source Estimation Based on Joint Source-Filter Model Optimisation. // IEEE Trans. Audio, and Speech Process. 2006. Vol. 14. № 2. P. 492–501.



Макаров И.С.

Об одном алгоритме оценки формантных частот на интервале сомкнутых голосовых складок

16. *Леонов А.С., Сорокин В.Н., Макаров И.С.* Устойчивость оценок формантных частот. // Речевые технологии. 2009. Т. 1. С. 3–21.
17. *Akande O., Murphy P.* Estimation of the Vocal Tract Transfer Function with Application to Glottal Wave Analysis. // *Speech Communication*. 2005. Vol. 46. № 1. P. 15–36.
18. *Nordstrom K., Tzanetakis G. and Driesse P.* Transforming Perceived Vocal Effort and Breathiness Using Adaptive Pre-emphasis Linear Prediction. // *IEEE Trans. Audio, Speech, and Language Process.* 2008. Vol. 16. № 6. P. 1087–1096.
19. *Miyoshi Y., Yamato K., Mizoguchi R., Yanagida M. and Kakusho O.* Analysis of Speech signals of Short Pitch Period by a Sample-Selective Linear Prediction. // *IEEE Trans. Acoust., Speech, and Signal Process.* 1987. Vol. 35. № 9. P. 1233–1240.
20. *Ma Ch., Kamp Y. and Willems L.* Robust Signal Selection for Linear Prediction Analysis of Voiced Speech. // *Speech Communication*. 1993. Vol. 12. P. 69–81.
21. *Magi C., Pohjalainen J., Backstrom T., and Alku P.* Stabilised Weighted Linear Prediction. // *Speech Communication*. 2009. Vol. 51. P. 401–411.
22. *Yegnanarayana B., Veldhuis R.* Extraction of Vocal-Tract System Characteristics from Speech Signals. // *IEEE Trans. Speech, and Audio Process.* 1998. Vol. 6. № 4. P. 313–327.
23. *Stoica P., Soderstrom T. and Ti F.* Asymptotic Properties of the High-Order Yule-Walker Estimates of Sinusoidal Frequencies. // *IEEE Trans. Acoust., Speech, and Signal Process.* 1989. Vol. 37. № 11. P. 1721–1734.
24. *Cranen B., Boves L.* On Subglottal Formant Analysis. // *J. Acoust. Soc. Amer.* 1987. Vol. 81. № 3. P. 734–746.
25. *Sreenivas T., Niederjohn R.* Zero-Crossing Based Spectral Analysis and SVD Spectral Analysis for Formant Frequency Estimation in Noise. // *IEEE Trans. Signal Process.* 1992. Vol. 40. № 2. P. 282–293.
26. *Kot A., Tufts D. and Vaccaro R.* Analysis of Linear Prediction by Matrix Approximation. // *IEEE Trans. Signal Process.* 1993. Vol. 41. № 11. P. 3174–3177.
27. *Тихонов А.Н., Арсенин В.Я.* Методы решения некорректных задач. М.: Наука, 1986.
28. *Ekman A., Kleijn B. and Murthi M.* Regularized Linear Prediction of Speech. // *IEEE Trans. Audio, Speech, and Language Process.* 2008. Vol. 16. № 1. P. 65–75.
29. *Беллман Р.* Динамическое программирование. М.: Изд-во иностр. лит-ры. 1960.
30. *Лоусон К., Хенсон Р.* Численное решение задач метода наименьших квадратов. М.: Наука, 1986.
31. *Цыплихин А.И.* Анализ импульсов голосового источника. // *Акустический журнал*. 2007. Т. 53. № 1. С. 119–133.
32. *Hawks J.* Difference Limens for Formant Patterns of Vowel Sounds. // *J. Acoust. Soc. Amer.* 1994. Vol. 95. P. 1074–1084.
33. *Westbury J.* *X-ray Microbeam Speech Production Database. User's Handbook. Version 1.* — 1994.

Сведения об авторе

Макаров Илья Сергеевич —

кандидат технических наук, заведующий лабораторией анализа и обработки биомедицинских сигналов ОАО НПО «Экран».

Старший научный сотрудник лаборатории когнитивных интерфейсов Российского научного центра «Курчатовский институт».

Действительный член Академии медико-технических наук РФ.

Научные интересы — биологическая и медицинская акустика, акустика речи.

Имеет более 40 публикаций.

Email: speechprod_mak@mail.ru