OPTIMAL SPEECH SIGNAL PARTITION INTO ONE-QUASIPERIODICAL SEGMENTS

Taras K. Vintsiuk

Speech Science and Technology Department
International Research-Training Centre for Information Technologies and Systems
40 Prospekt Akademika Hlushkova, Kyjiv 03680 Ukraine
vintsiuk@uasoiro.org.ua

ABSTRACT
Quasi-periodicity and non-periodicity signal models are proposed. Each hypothetical one-quasiperiodical signal segment is considered as a random distortion of previous or following one taken with the unknown multiplying factor. The problem of optimal current pitch period discrimination and speech signal partition into one-quasiperiodical microsegments consists in the finding of the best one-quasiperiod beginnings or the one-quasiperiodical segments under restrictions on both value and changing of current quasiperiod duration and multiplying factor. For this problem solving an effective algorithm based on dynamic programming is proposed and its application for speech signal analysis, recognition and synthesis is discussed.

1. INTRODUCTION
It is well known that an analysis of such complicated signals as speech signals has to be carried out synchronously with a current pitch period (quasiperiod). Besides for a speech signal it is important to pick out large quasi-periodic and non-periodic (noise) segments and to find current one-quasiperiodical segment durations, beginnings and ends as well.

2. QUASI-PERIODICITY AND NON-PERIODICITY MODELS.
PROBLEM STATEMENT
Let the signal $f_n, n = 1:N$ be observed where $f_n$ is a signal value at the discrete uniform time $n\Delta t$ with step $\Delta t$, for example $\Delta t = 50\mu s$ for speech signal. If the $(s+1)$-th one-quasiperiod signal segment beginning is denoted by $n_s$, then $(n_{s+1}-1)$ will be the end of the $s$-th one-quasiperiod. Further segment signal

$$f_{n_{s+1}+j}, j = 0:(T_s - 1), T_s = n_s - n_{s-1} \quad (1)$$

will be called $s$-th one-quasiperiod segment with duration $T_s$, if it is approximated sufficiently well by neighbour ones, $(s-1)$-th or $(s+1)$-th, which respectively precedes or follows the $s$-th one-quasiperiod segment. The latter are taken with the unknown multiplying number $\alpha_s$ or $\alpha_s^+$:

$$f_{n_{s-1}+j}^- = \begin{cases} \alpha_s^- f_{n_{s-1}+j}^-, & j = 0: (\min(T_s, T_{s+1}) - 1), \\ 0, & j = \min(T_s, T_{s+1}): (T_s - 1), \end{cases} \quad (2)$$

$$f_{n_{s+1}+j}^+ = \begin{cases} \alpha_s^+ f_{n_{s+1}+j}^+, & j = 0: (\min(T_s, T_{s+1}) - 1), \\ 0, & j = \min(T_s, T_{s+1}): (T_s - 1). \end{cases} \quad (3)$$

Let us introduce a priori restrictions for multiplying number value $\alpha$ and both current quasiperiod duration value $T_s$ and its changing $\Delta_s = T_s - T_{s-1}$:

$$\alpha > 0, \quad T_{\min} \leq T_s \leq T_{\max}, \quad |\Delta_s| \leq \Delta_{\max}. \quad (4)-(6)$$

Now let us fix the elementary quasi-periodicity (EQP) measure for the $s$-th one-quasiperiodical signal segment $f_{n_{s-1}+j} = f_{n_s-T_s+j}, j = 0:(T_s - 1)$ as

$$d^-(n_s, T_s, \Delta_s) = \min_{\alpha > 0} \sum_{j=0}^{T_s-1} (f_{n_{s-1}+j} - f_{n_{s+1}+j})^2 =$$
$$= \min_{\alpha > 0} \sum_{j=0}^{\min(T_s, T_{s+1})-1} (f_{n_{s-1}+j} - \alpha f_{n_{s-2}+j+\Delta_s})^2 + \sum_{j=\min(T_s, T_{s+1})}^{T_s-1} f_{n_{s-2}+j+k}^2 \quad (7)$$

or as

$$d^+(n_s, T_s, \Delta_{s+1}) = \min_{\alpha > 0} \sum_{j=0}^{T_s-1} (f_{n_{s+1}+j} - f_{n_{s-1}+j})^2 =$$
$$= \min_{\alpha > 0} \sum_{j=0}^{\min(T_s, T_{s+1})-1} (f_{n_{s+1}+j} - \alpha f_{n_{s-1}+j})^2 + \sum_{j=\min(T_s, T_{s+1})}^{T_s-1} f_{n_{s+2}+j+k}^2 \quad (8)$$

As it is followed from the expression (7) and (8) the signal segment $f_{n_{s-1}+j} = f_{n_s-T_s+j}, j = 0:(T_s - 1)$ is tested on quasi-periodicity by comparison with previous segment $f_{n_{s-2}+T_s+\Delta_s+j}, j = 0:(T_s - \Delta_s - 1)$ or
following one \( f_{n_{s}+j}, \ j = 0 : (T_{s} + \Delta s_{+1} - 1) \) but only the best comparison result will be associated with the quasiperiodicity measure value EQP.

The EQP measure can be simplified. Let us omit the index \( s \) and denote:

\[
E(n) = \sum_{j=1}^{T_{n}} f_{j}^{2}, \tag{9}
\]

\[
D(n, T) = \sum_{j=0}^{T_{n}} f_{n-T+j}^{2} = E(n) - E(n - T), \tag{10}
\]

\[
S^{-}((n, T), \Delta) = \sum_{j=0}^{\min(T-\Delta T)-1} f_{n-T-j} f_{n-2T+\Delta j}, \tag{11}
\]

\[
M^{-}((n, T), \Delta) = \sum_{j=0}^{\min(T-\Delta T)-1} f_{n-2T+\Delta j}^{2} = \tag{12}
\]

\[
= E(n - \max(T - \Delta, T)) - E(n - 2T + \Delta), \tag{13}
\]

\[
S^{+}((n, T), \Delta) = \sum_{j=0}^{\min(T+\Delta T)-1} f_{n-T-j} f_{n+j}, \tag{14}
\]

\[
M^{+}((n, T), \Delta) = \sum_{j=0}^{\min(T+\Delta T)-1} f_{n+j}^{2} = \tag{15}
\]

\[
= E(n + \min(T + \Delta, T)) - E(n), \tag{16}
\]

\[
\alpha^{\pm}((n, T), \Delta) = \frac{S^{+}((n, T), \Delta)}{M^{+}((n, T), \Delta)}. \tag{17}
\]

\[
\beta^{\pm}((n, T), \Delta) = \alpha^{\pm}((n, T), \Delta) \left\{ \begin{array}{ll}
0, & \text{if } \alpha^{\pm}((n, T), \Delta) \leq 0; \\
\alpha^{\pm}((n, T), \Delta) S^{-}((n, T), \Delta), & \text{if } \alpha^{\pm}((n, T), \Delta) > 0;
\end{array} \right. \tag{18}
\]

Now let us rewrite the EQP measure as

\[
d^{\pm}((n, T), \Delta) = D(n, T) - \beta^{\pm}((n, T), \Delta). \tag{19}
\]

Any permissible variant \( (n_{s}, T_{s}) \), \( s = 0, 1, 2, \ldots, P \) of the signal \( f_{n}, n = 1 : N \) segmentation on \( P \) one-quasiperiodic segments under restrictions (4)—(6) is characterized by the sum of respective EQP measure values:

\[
G(n_{s}, s = 0 : P) = \sum_{s=0}^{P} d^{\pm}((n_{s}, T_{s}), \Delta) = \tag{20}
\]

\[
= \sum_{s=0}^{P} \left( D(n_{s}, T_{s}) - \beta^{\pm}((n_{s}, T_{s}), \Delta) \right) = \tag{21}
\]

\[
= E(N + 1) - \sum_{s=0}^{P} \beta^{\pm}((n_{s}, T_{s}), \Delta). \tag{22}
\]

To find for the signal \( f_{n}, n = 1 : N \) the best segmentation onto unknown number \( P \) one-quasiperiodic segments it is necessary to minimize the criteria (18) or, this is the same, to maximize the criteria (19) on all permissible sequences \( n_{s}, s = 0 : P \):

\[
H(n_{s}, s = 0 : P) = \sum_{s=0}^{P} \beta^{\pm}((n_{s}, T_{s}), \Delta). \tag{23}
\]

Let us solve the problem (17)-(18) or (19) of optimal speech signal \( f_{n}, n = 1 : N \) partition into one-quasiperiodic segments under restrictions (4)—(6) on quasiperiodic intensity, duration and duration changing. Then, at the second stage, we will unite one-quasiperiodic segments into large quasiperiodic or non-periodic ones accordingly with the relative approximation error for each optimal one-quasiperiodic segment.

**3. OPTIMAL SIGNAL PARTITION INTO ONE-QUASIPERIOD SEGMENTS**

The optimal segmentation problem (17)-(18) or (19) is being solved by dynamic programming method because these criteria satisfy the Bellman principle. They are the sums of elementary quasiperiodicity measure values \( d^{\pm}((n_{s}, T_{s}), \Delta), s = 0 : P \) for separate one-quasiperiod segment with number \( s \) and what’s more EQP measure takes into account the restrictions (5)—(6) on one-quasiperiod duration \( T_{s} \) and its changing \( \Delta \).

Let us denote by \( \Omega^{-}((u, T)) \) the set of all possible segmentations for the signal \( f_{i}, i = 1 : u, u < N \) into one-quasiperiodic segments under restrictions (5)—(6) and of what else that last one-quasiperiod begins at the time \( u - T \), ends at the time \( u - 1 \), so has the duration \( T_{s} \) and is compared with last past one one-quasiperiods accordingly with “−” EQP measure. Let \( H^{-}((u, T)) \) be the optimal segmentation criterion value for the signal \( f_{i}, i = 1 : u, u < N \) which is reached on the segmentation variants set \( \Omega^{-}((u, T)) \). Let \( \Delta(u, T) \) be the best changing for the last one-quasiperiod duration in the optimal segmentation variant from \( \Omega^{-}((u, T)) \). Then the \( \beta^{-}((u, T), \Delta(u, T)) \) is the last one-quasiperiodic contribution to the sum \( H^{-}((u, T)) \).

Similarly, let us introduce the set \( \Omega^{+}((u, T)) \) of all possible segmentations for the reduced signal \( f_{i}, i = 1 : (u - T) \). This set is considered in order to have prepared to compare the following one-quasiperiod \( f_{i}, i = (u - T) : u \), accordingly with “+” EQP measure. Then for the analyzed signal \( f_{i}, i = 1 : u, u < N \) the quantity \( H^{+}((u, T)) \) will be the optimal segmentation criterion value on the set \( \Omega^{+}((u, T)) \).

Let the values \( H^{-}((u, T)) \) and \( \Delta(u, T) \)，and \( H^{+}((u, T), \Delta) \) are already calculated for all times \( u = 1 : (n - 1) \), all possible values \( T = T_{\min} : T_{\max} \) and all
\[ |\Delta| \leq \Delta_{\text{max}}. \]

Then for the initial signal subsequence \( f_i \), \( i = 1:n \) with time duration \( n \) the potentially optimal values \( H^-(u, T), \Delta(n, T), H^+(u, T), \Delta \) for all \( T = T_{\text{max}} \) \( T_{\text{max}} \) and all \( |\Delta| < \Delta_{\text{max}} \) can be calculated by the next recurrent formulae:

\[
H^-(n, T) = \max_{|\Delta| < \Delta_{\text{max}}} \left\{ \max \left[ H^-((n - T), T - \Delta) \right], \right. \\
H^+(n - T, T - \Delta) + \beta^+((n - T, T - \Delta), \Delta) \right. \\
\left. + \beta^+((n, T), \Delta) \right\}; \quad \text{(20)}
\]

\[
(\Delta^-, \bar{\Delta})(n, T) = \arg\max_{|\Delta| < \Delta_{\text{max}}} \left\{ \max \left[ H^-((n - T), T - \Delta) \right], \\
H^+(n - T, T - \Delta) + \beta^+((n - T, T - \Delta), \Delta) \right. \\
\left. + \beta^+((n, T), \Delta) \right\}; \quad \text{(21)}
\]

\[
H^+(n, T) = \max_{|\Delta| < \Delta_{\text{max}}} \left\{ \max \left[ H^-((n - T), T - \Delta) \right], \\
H^+(n - T, T - \Delta) + \beta^+((n - T, T - \Delta), \Delta) \right. \\
\left. + \beta^+((n, T), \Delta) \right\}; \quad \text{(22)}
\]

In formulae (21), (23) records \( (\Delta, \bar{\Delta})(n, T) \) acquire values \( (\Delta, \gamma) \) or \( (\Delta, \alpha) \) respectively depending on whether first or second expression in \( \max \) is better.

It is possible to specify all variants of speech signal segmentation by graph the nodes of which are determined by the pair of co-ordinates \( (n, T) \). But in reality each node \( (n, T) \) consists of two subnodes \( \bar{n}(n, T) \) and \( * (n, T) \). To these subnodes the potentially optimal segmentation criterion values and the potentially optimal one-quasiperiod duration changing are imputed:

\[
H^-(n - T, T - \Delta), (\Delta^-, \bar{\Delta})(n, T) \quad \text{and} \quad H^+(n - T, T - \Delta), (\Delta^+, \bar{\Delta})(n, T)
\]

respectively. Two subnodes \( \bar{n}(n, T) \) and two subnodes \( \bar{n}(n, T) \) where \( T_{\text{max}} \leq T \leq T_{\text{max}} \) and \( |\Delta| \leq \Delta_{\text{max}} \), are connected by the arrows that enter to the subnodes \( \bar{n}(n, T) \), and to these four arrows the EQP measure values \( \beta^-((n, T), \Delta), \beta^+((n, T), (n - T, T - \Delta) \Delta) \) \( + \beta^+((n, T), \Delta) \) are prescribed, respectively. Permissible arrows are associated so that the final node of first arrow is the start node of second one. These associated arrows determine permissible variant (trajectory) of signal segmentation onto one-quasiperiodic segments. The sum of EQP measure values prescribed to arrows of permissible segmentation trajectory characterizes the quality of this segmentation trajectory. Thus the signal partition of speech signal onto one-quasiperiodic segments is reduced to the finding the best trajectory on the graph.

To realize the calculations (20)-(21) it is necessary to specify both the initial \( W_{\text{start}} \) and final \( W_{\text{final}} \) sets of nodes, and the respective values \( H(n, T) \) and \( \Delta(n, T) \) for \( (n, T) \in W_{\text{start}} \). Let us take into account that \( f_i \neq 0 \) for all \( n \leq 1 \) and all \( n > N \).

Further there will be given some important statements.

Let \( W(n) \) will be the minimal set of nodes (MNS) with the position \( n \):

\[
W^-(n) = \{(v, T) : T = T_{\text{max}}, n = (n - T): (n - 1)\}, \quad \text{(24)}
\]

\[
W^+(n) = \{(v, T) : T = T_{\text{max}}, n = (n + 1) : (n + T)\}. \quad \text{(25)}
\]

Statement 1. All permissible segmentation trajectories penetrate all MNS \( W^-(n) \) with the position \( n \).

Statement 2. The initial set \( W_{\text{start}} \) is specified by \( W^-(1) \).

Statement 3. The value \( H(n, T) \) and \( \Delta(n, T) \) for \( (n, T) \in W_{\text{start}} \) is \( W^-(n) \) are determined as

\[
H^-(n, T) = 0, \quad (\Delta^-, \bar{\Delta})(n, T) = 0. \quad \text{(26)-(27)}
\]

Statement 4. The final set \( W_{\text{final}} \) of nodes is specified as \( W^+(N+T_{\text{max}}) \).

The results of recurrent speech signal processing are given by the array \( (\Delta^+, \bar{\Delta})(\mathbf{n}, T) \) calculated with (21) and (23) for all \( n = 2(N+T_{\text{max}}) \).

Optimal one-quasiperiodical segmentation forming is based on the calculated array \( (\Delta^+, \gamma)(n, T) \) where \( \delta \) and \( \gamma \) become values \( \bar{\gamma} \) or \( \gamma \) independently. On the node set \( W_{\text{final}} = W^+(N+T_{\text{max}}) \) we find out \( H^+ \) and \( \mathbf{n} \), \( \mathbf{T} \):

\[
H^+ = \max_{(n, T) \in W_{\text{final}}} \max \left[ H^-(n, T), H^+(n, T) \right]. \quad \text{(28)}
\]

Then, further, using the recurrent extracting formulae (30) (starting from the end of signal), let make notes the one-quasiperiodic beginnings and its durations:

\[
n^*_0 = n^*, \quad T^*_0 = T^*, \quad \delta^*_0 = \bar{\gamma}, \quad n^*_s \neq T^*_s \leq 1, \quad s = 1,2, ..., (n^*_1, T^*_1): \quad n^*_1 = n^*_{s-1} - T^*_s-1, \quad T^*_1 = T^*_s-1 - \Delta^s \delta^*_s (n^*_{s-1}, T^*_{s-1}), \quad \delta^*_s = \gamma \text{ from } \left(\Delta^s \delta^*_s, \bar{\gamma}\right)(n^*_{s-1}, T^*_{s-1}), \quad P^* = P^* + 1 \text{ until } n^*_s < 1. \quad \text{(30)}
\]

The \( H^+ \) is the optimal value for the criterion (19). The \( n^* \) is the beginning for the last one-quasiperiod segment with the duration \( T^*_0 \), the \( n^*_1 \) is the beginning for the last one one-quasiperiod segment with the duration \( T^*_1 \) and so on. The \( P^* \) is the optimal number of quasiperiods.
Statement 5. The calculations for $\beta^*((n, T), \Delta)$ accordingly to (9)-(16) are being essentially simplified if there are used the next recurrent expressions:

\[
E(n) = E(n-1) + f_{n-1}^2, \quad (32)
\]

\[
S^*(((n-1), T, 0) - f_{n-i-T} f_{n-1} + f_{n-i-T+1}, \quad (33)
\]

\[
S^*((n, T), \Delta) = S^*((n, T), \Delta - 1) - f_{n-\Delta-T} f_{n-\Delta}, \quad \text{for } \Delta > 0, \quad (34)
\]

\[
S^*((n, T), \Delta) = S^*((n-T, T-\Delta), \Delta). \quad (35)
\]

At the final stage the found one-quasiperiodic segments are united into large quasiperiodic or non-periodic segments. For this the relative approximation error for the neighbouring one-quasiperiod microsegments, a priori information about voiced/unvoiced segments duration and so on are taken into account.

**4. EXAMPLE**

In Fig.1 there are shown the results of speech signal partition into one-quasiperiod segments. The numbers in columns mean current one-quasiperiod duration $T^*$, duration changing $\Delta^*$, analyzed one-quasiperiod segment is approximated by previous $\Rightarrow$ or following $\Leftarrow$ one, one-quasiperiod energy $E^*$, criterion value $H^*$, EQP measure value $\beta^*$, one-quasiperiod beginning $n^*$, multiplying factor value $\alpha^*$, one-quasiperiod approximation error energy $e^*$ and one-quasiperiod relative approximation error $r^*$ respectively.

**5. CONCLUSION**

The proposed procedure for optimal partition into one-quasiperiod segments is applied to any kind of speech signal, whenever it is voiced or unvoiced. So, it gives one-quasiperiodical interpretation for noised as well as for voiced sounds.

This procedure is designed for appliance in automatic speech analysis, recognition and synthesis, especially if they are doing in time domain. For example, a subsequence of some neighboring one-quasiperiod segments can be transmitted as a first one-quasiperiod segment signal repeated several times with respective one-quasiperiod durations and multiplying factors (time vocoder). As for speech recognition, the picked out one-quasiperiod segments with its one-quasiperiod durations and multiplying factors form speech elements sequence to be recognized. In speech synthesis to control pitch, it is used one-quasiperiod reduction or one-quasiperiod covariation prediction of memorized one-quasiperiod speech microelements.