SCIENTIFIC PAPERS
OF THE UNIVERSITY OF PARDUBICE
Series B
The Jan Perner Transport Faculty
12 (2006)

NEW METHODS PROPOSAL FOR VoIP TRANSMISSION

Katarína BACHRATÁ, Juraj BOROŇ, Boris ROSIVAL

Katedra informačných sietí, Fakulta riadenia a informatiky, Žilinská Univerzita, Slovensko

1. Introduction and preliminaries

The IP telephony service uses for the speech transmission [1], [4] the principle of the packet commutation. Transmission of the speech signal through the data network requires conversion of the analog to the discrete signal. Mathematical model of this conversion is decomposition of the original signal to the generalized Fourier basis:

$$f(t) = \sum_{k=-\infty}^{\infty} c_k \Phi_k(t)$$

In the case of the orthogonal basis the coefficients c_k are computed as a ratio of inner products of original signal and related basis functions:

$$c_k = \frac{\left\langle f(t), \Phi_k(t) \right\rangle}{\left\langle \Phi_k(t), \Phi_k(t) \right\rangle}, \quad \text{for} \quad k = 1, 2, 3\ldots$$

The Shannon basis used to be the best for speech signal conversion. There were two serious reasons for using the Shannon basis. First: the computing of the coefficients c_k is very simple, because c_k are equal to samples $f(k\Delta)$ of the original signal $f(t)$. According to the Nyquist–Shannon sampling theorem:

The exact reconstruction of a continuous signal from its samples is possible if the signal is bandlimited and the sampling frequency is greater than twice the signal bandwidth.↑

A mathematical expression of this theorem is

$$f(t) = \sum_{k=-\infty}^{\infty} c_k \Phi_k(t) = \sum_{k=-\infty}^{\infty} f(k\Delta)\Phi(t-k\Delta) = \sum_{k=-\infty}^{\infty} f(k\Delta)\frac{\sin\Omega(t-k\Delta)}{\Omega(t-k\Delta)}, \quad \Omega = \frac{\pi}{\Delta}$$

The second reason is in the properties of the basis ortogonality. Ortogonal basis allows the most effective transmission of the speech signal. The information in each coefficient is clear, non-affected by the other coefficients. In the figure 1 we can see the shape of the generating basis function

$$\Phi(t) = \frac{\sin \Omega t}{\Omega t}$$

and five elements of Shannon basis created by this function.

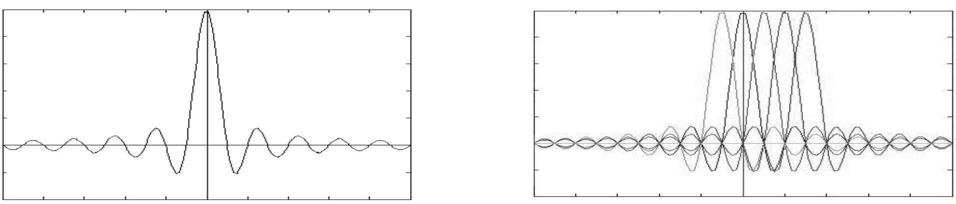


Fig. 1 *One and five elements of the Shannon basis*

The values of the $\Phi(t)$ are 1 in $t = 0$ and 0 in $t = k.\Delta$ for $k = 1,2,\ldots$.

In figure 2 we can see the reconstruction of the signal using one, three and then eight elements of the Shannon basis. Evidently, the coefficient of one element does not impact the other elements of basis. In addition, the approximation is better for the values near to used coefficients and vanished in bigger distances.

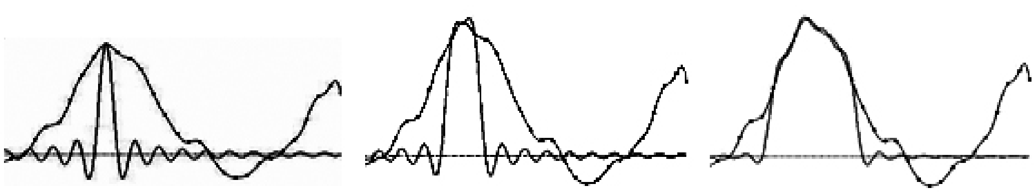


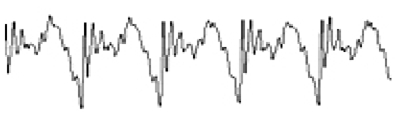
Fig. 2 *The original signal and its approximation by one; three and eight elements of the Shannon basis*

Nowadays for the transmission speech signal through the IP networks is used decomposition to the Shannon basis. We must not keep the benefits of the Shannon decomposition as: easy computing of coefficients and orthogonality of the basis. More complicated computing of coefficients can be realized in the sufficient short time because of the power of computer equipments. The orthogonality of basis can be explained not only as a clear representation of the signal, cleaned from redundant information. It can be understood as a transformation with non correlated coefficients. During the transmission over IP network, packets may delay, due to the jitter of network and consequently get lost. The loss of uncorrelated coefficient causes the loss of whole information contained in lost basis element.

The aim of our work is to find and test the new basis for speech decomposition. This new basis keep the good properties of Shannon basis. Useless properties are replaced by correlation between coefficients and following useful possibility to partly reconstruct the missing parts of original signal.

2. Conditions for basis construction

However the common conception for stochastic modeling of a speech process is the stationary random process. We present the results achieved by modeling short segments of speech as a cyclostationary stochastic process. This model corresponds with a way of creation of the speech in a human vocal tract. The vibration of the vocal chords and resonance in the vocal sinuses causes that the output sound (especially sound of vowels) has rather cyclic or periodical character. We can see this character in figure 3, where is the mean and the correlation of the real sound of vowel ‘E’, which takes about 20 ms.



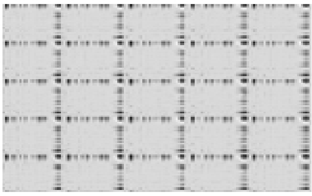


Fig. 3 *The real sound of vowel ‘E’ running 20 ms. The mean as a function of time and correlation as a two dimensional function of time variables.*

Let us denote time variable as t and stochastic variable as ξ. Stochastic process is called cyclostationary if its mean and correlation have a periodical character:

$$E\{f(t+\Delta,\xi)\} = E\{f(t,\xi)\} \quad \text{and} \quad R(t_1 + m\Delta, t_2 + m\Delta) = R(t_1, t_2)$$

for all $m \in Z$

The theorem below is shown in [9]: Let $f(t,\xi)$ be a stochastic process

$$f(t,\xi) = \sum_{n=-\infty}^{\infty} c_n g(t - n\Delta)$$

with deterministic $g(t)$, and stationary sequence of random variables c_n. Then the process $f(t,\xi)$ is cyclostationary process.↑

The theorem explains that the sufficient condition of cyclostationarity is that the basis used for discretization is generated of regularly shifted function. This property is satisfied for instance for the Shannon base, shift invariant spaces and in a special wavelet subspaces. Using two ways (observation the real voice and investigation the stochastic properties) we achieve the same condition, cyclostationarity of the voice process.

The condition of cyclostationarity can be obtained if the basis is generated by shifting of one function. Unfortunately, often, the set of shifted functions is not orthogonal. The Shannon basis is orthogonal and the coefficients can be computed as the coefficients in the generalized Fourier series.

If the orthogonality is not satisfied the unique decomposition must be done by another way. Instead of orthogonality they use the independence of the basis elements. The linear independence of basis is defined only for finite-dimensional spaces, but it is necessary to use the infinite-dimensional spaces for the expression of voice signal. There can be used the ω-independence of the basis in the infinite-dimensional spaces:

A system of the functions $\{\Phi_k\} \subset L_2(-\infty,\infty)$ is called ω-independent if and only if

$$\sum_{k=-\infty}^{\infty} c_k \Phi_k = \theta \qquad \Rightarrow \qquad c_k = 0, \quad \forall k \in Z$$

The condition for the generating functions $g(t)$ helps to verify the ω-independence of the basis:

Let $\Delta > 0$ and function $g \in L_2(-\infty,\infty)$. The function g generates the ω-independent system $\{\Phi_k\} = \{g(t - k\Delta)\}$ if and only if

$$\sum_{n=-\infty}^{\infty} \left| \hat{\Phi}\left(\omega + \frac{2\pi n}{\Delta} \right) \right| \neq 0 \qquad for \qquad \omega \in \left\langle -\frac{\pi}{\Delta}, \frac{\pi}{\Delta} \right\rangle \ a.e.$$

Where $\hat{\Phi}(\omega)$ is a Fourier transform of the function $g(t)$: $\hat{\Phi}(\omega) = \int_{-\infty}^{\infty} g(t) e^{-j\omega t} dt$.

Different shapes of the function $g(t)$ lead to different properties of the original signal decomposition. An important property in IP telephony is the correlation between the coefficients of decomposition. We can achieve the correlation between the coefficients by using the function $g(t)$ which shifts intersect each other. In future (for real transmission) we will use the function $g(t)$ with support about 100 Δ and there will be 200 intersections.

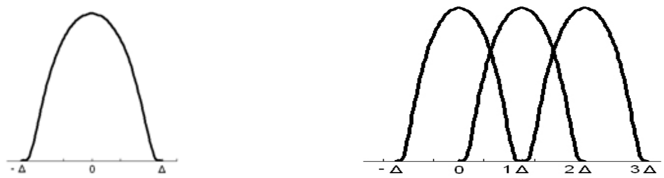


Fig. 5 *The generating function* $g(t) = e^{-\frac{1}{(t-1)(1-t)}}$ *and its shift with* $\Delta = \dfrac{1}{2}$

3. Calculation of basis coefficients

The calculation of coefficients c_m in the decomposition $f(t) = \sum_{m=-\infty}^{\infty} c_m \Phi_m(t)$ gets the system of the linear equations:

$$\langle f(t), \Phi_n(t) \rangle = \sum_{m=-\infty}^{\infty} c_m \langle \Phi_m(t), \Phi_n(t) \rangle, \qquad for \quad k = 1, 2, \dots$$

Where:

$$\langle f(t), \Phi_n(t) \rangle = \int_{-\infty}^{\infty} f(t) g(t - n\Delta) dt$$

$$\langle \Phi_m(t), \Phi_n(t) \rangle = \int_{-\infty}^{\infty} g(t - m\Delta) g(t - n\Delta) dt \quad \forall m, n \in Z$$

When substitute

$$f(t) = \sum_{k=-\infty}^{\infty} f(k\Delta) \frac{\sin \Omega(t - k\Delta)}{\Omega(t - k\Delta)}$$

$$\langle f(t), \Phi_n(t) \rangle = \left\langle \sum_{k=-\infty}^{\infty} f(k\Delta) \frac{\sin \Omega(t - k\Delta)}{\Omega(t - k\Delta)}, \Phi_n(t) \right\rangle = \sum_{m=-\infty}^{\infty} c_m \langle \Phi_m(t), \Phi_n(t) \rangle, \qquad for \quad k = 1, 2, \dots$$

$$\sum_{k=-\infty}^{\infty} f(k\Delta) \left\langle \frac{\sin \Omega(t - k\Delta)}{\Omega(t - k\Delta)}, \Phi_n(t) \right\rangle = \sum_{m=-\infty}^{\infty} c_m \langle \Phi_m(t), \Phi_n(t) \rangle, \qquad for \quad k = 1, 2, \dots$$

Let us denote

$S_{kn} = \int_{-\infty}^{\infty} \frac{\sin \Omega(t-k\Delta)}{\Omega(t-k\Delta)} g(t-n\Delta)dt$ and $B_{mn} = \int_{-\infty}^{\infty} g(t-m\Delta)g(t-n\Delta)dt$, then we get for coefficients c_m:

$$\sum_{k=-\infty}^{\infty} f(k\Delta)S_{kn} = \sum_{m=-\infty}^{\infty} c_m B_{mn}, \qquad for \quad n = 1, 2, \ldots$$

The coefficients S_{kn} and B_{mn} can be computed offline and saved into the database. Computation in real time is required for the solving of system linear equations. Therefore the power of the generating function must be finite. We can choose the function $g(t)$ with finite support. Hence the constants $B_{mn} = 0$ for all m, n such $|(m-n)\Delta| > \varepsilon$ and the sparse systems of linear equations has been created. There are special methods for solving sparse systems, see [10].

4. Impact of the coefficients loss to the voice quality

Quality of the IP telephony service can be measured by MOS (mean opinion score), which gives the subjective point of view. The technical parameters of the quality bring the more objective information. Signal-to-noise ratio is one of such technical parameters. It is the ratio of the power of transmitted signal and the noise signal (difference between transmitted signal and received signal).

$$SNR = 10 \log \frac{\sigma_f^2(t)}{\sigma_n^2(t)} [dB]$$

During the transmission over IP network, packets may get delayed, or lost. For transmission we can use basis of function $g(t)$ with finite support. The theory of the impact of packet loss to the signal-to-noise ratio in cases of the shifts of such function $g(t)$ has been elaborated in article [8]. The shape of the generating function affects to SNR. If the lost coefficients are substitute by zero there can be find the limitations for SNR, see figure 4. The basis B_m is created by the function $g(t)$, which support has m nonzero intersections with supports of the shifts $g(t-n\Delta)$ .The correlation between coefficients increases with the number of intersections of the supports of basis elements.

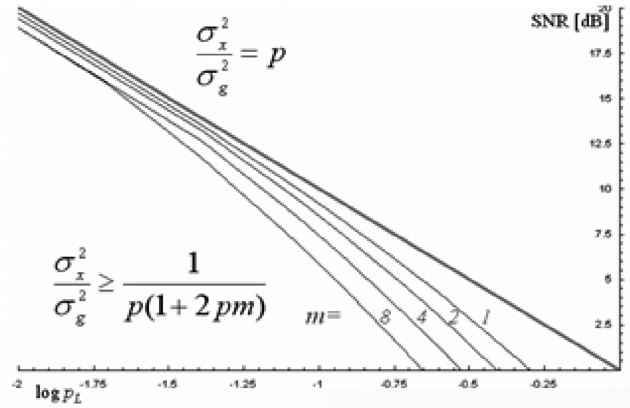


Fig. 5 *Limitation for SNR in case of basis B_1, B_2, B_4, B_8*

5. Results of simulations

The first experiment with the simulation of the loss of packets rests in the decomposition into the classical Fourier basis. We can see the impact of the loss of Shannon samples and Fourier coefficients to the SNR in the figure 6.

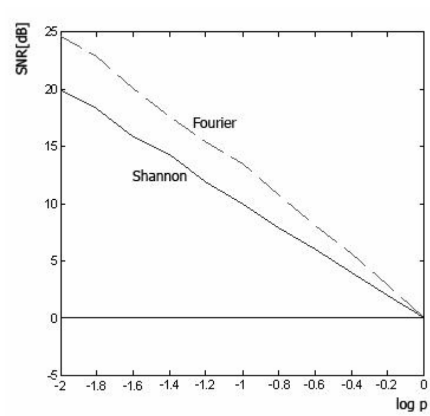


Fig. 6 *Results of the simulation for the SNR in case of Shannon and Fourier decomposition.*

The impact of coefficients loss shows, that the Fourier decomposition could be better, than Shannon. But for such decomposition is necessary to have quite a long segment of the speech signal. There is not time enough for use long segments. So the Fourier basis is not very useful.

The example below shows the results of the decomposition into basis with 'hat' generating function $g(t) = e^{-\frac{\Delta^2}{(t-\Delta)(\Delta-t)}}$ and its shifts $g(t-n\Delta)$.

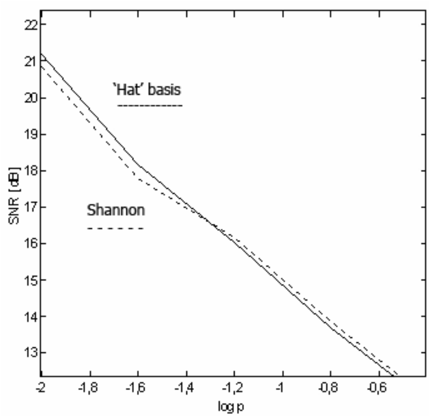


***Fig. 7** Results of the simulation for the SNR in case of 'hat' decomposition.*

The comparison of the results of SNR for the simulation of correlated coefficients loss and the Shannon samples loss can be found in figure 7:

6. Conclusions

The problem how to save the quality of voice signal is discussed in [4]. The quality can be measured by using the estimations for noise signal. One of the criterions of quality is the signal to noise ratio. We have showed in this article, that the decomposition into non-Shannon bases can bring better results than Shannon decomposition. Actually, the theoretical limitations in this paper visualizes, that the results can be a little worse, but in well-advised case of the choice for the shape of generating function, results can be much better. The results of our first experiment with non orthogonal basis give the expectations of the useful using of this method. The quality of the transmission has to be check by MOS. It will be aim of the next research.

Literature

1. KLIMO, M., KOVÁČIKOVÁ, T., SEGEČ, P. *Hlas cez Internet.* Košice, Elfa, 2005, ISBN 80-8086-007-6
2. W. CHEN, S. ITOH, and J. SHIKI *On sampling in shift invariant spaces,* IEEE Trans. Information Theory, vol. 48, no. 10, pp. 2802-2810, 2002.
3. W. A. GARDNER, L. E. FRANKS: *Characterization of Cyclostationary Random Signal Processes*, IEEE Trans. on Information Theory, vol. 21, no. 1, pp. 4-14, 1975.

4. JAKAB, F,.ANDRÉ, M. *Measuring and Evaluation of QoS Parameters in IP Networks*, V.Doktorandská konferencia a študentská vedecká odborná súťaž TU v Košiciach, pp. 119-120, Košice, (2004), ISBN 80-968395-9-4
5. KLIMO, M.,BACHRATÁ, K. *Basis of Regularly Shifted Functions*, Scientific papers of the University of Pardubice, The Jan Perner Transport Faculty 9, (2003)
6. KLIMO, M.,BACHRATÁ, K. *Sampling of Function Generating Shift Invariant Spaces*, Scientific papers of the University of Pardubice, The Jan Perner Transport Faculty 10, (2004)
7. KLIMO, M.,BACHRATÁ, K. *Noise in Reconstruction of Speech Signal as Cyclostationary Stochastic Process*, Scientific papers of the University of Pardubice, The Jan Perner Transport Faculty 10, (2004)
8. KLIMO, M.,BACHRATÁ, K. *Impact of correlated loss on VoIP quality*, Scientific papers of the University of Pardubice, The Jan Perner Transport Faculty 11, (2005), ISSN 1211-6610, str.125-134
9. PAPOULIS, A. Probability, *Random Variables, and Stochastic Processes*, Polytechnic Institute of New York, McGraw-Hill, Inc., New York Toronto London, 1991
10. DAVIS, A., T. *Direct Methods for Sparse Linear Systems* SIAM, ISBN-10: 0-89871-613-6

Lektoroval: RNDr. Ludvík Prouza, CSc.

Předloženo: 20.3.2007

Resumé

NÁVRH NOVÝCH METÓD PRE PRENOS REČI CEZ IP SIETE

Katarína BACHRATÁ, Juraj BOROŇ, Boris ROSIVAL

Pre prenos zvukového signálu je potrebné urobiť konverziu analógového rečového signálu do digitálnej podoby. Pre klasickú telefóniu aj pre IP telefóniu sa používa rozklad do Shannonovej bázy. V článku sú zhrnuté doterajšie teoretické výsledky pre nové možnosti prenosu hlasu cez internet s použitím iných báz. Ich význam spočíva v možnosti lepšej rekonštrukcie pôvodného signálu aj v prípade, že niektoré hodnoty sa pri prenose stratia (resp. nie sú dodané včas a čakanie na ich príchod by spôsobilo degradáciu prijatého signálu). Závislosť medzi kvalitou rekonštruovaného signálu a percentom stratených paketov je odhadnutá pomocou odstupu signálu od šumu SNR. V článku sú popísané výsledky simulácie pre odhad SNR v prípade konkrétnej voľby generujúcej bázovej funkcie**.**

Summary

NEW METHODS PROPOSAL FOR VoIP TRANSMISSION

Katarína BACHRATÁ, Juraj BOROŇ, Boris ROSIVAL

For speech transmission is important to make conversion of analog speech signal to digital form. Signal conversion in classic and IP telephony uses Shannon basis decomposition. The Shannon basis used to be the best for speech signal conversion. The reason was, that the computing of the coeficients is very simple, because they are equal to samples of the original signal. In this paper are included teoretical results for new possibilities of transmission voice over IP using different basis. Their advantages rests in better reconstruction of the original signal also in case if we loose some of the data during the transmission(or they aren't delivered on time and waiting for them will cause degradation of received signal). The dependance between quality of

reconstructed signal and percentage of lost packets is estimated by signal to noise ratio SNR. In this paper are described results of simulation for SNR estimation in case of specific choice of generating basis function.

Zusammenfassung

DIE KONZEPTION FÜR NEUE METHODEN FÜR VoIP ÜBERTRAGUNG

Katarína BACHRATÁ, Juraj BOROŇ, Boris ROSIVAL

Wen wir wollen das Sprachsignal transferieren dazu brauchen wir die Konversion des analogisches Signales nach digitale Form.Für das klassisches telefonie und auch für IP telefonie wird die Abart nach Schannon Base benutzen. Im Artikel gibt es gleichzeitige teoretische Ergebnisse für das neuses möglichkeiten die Stimme durch das Internet transferieren mit anderen basen benutzung. Seine Bedeutung besteht in moglichkeit besser rekonstruktion des Originales Signales, auch im Fall das wir einige daten loswerden(bzw. die daten sind nicht rechtzetig übertragt was wurde das degeneration des Daten beibringen.) Abhängigkeit zwischen qualität des rekontruiert Signal und Prozent des Paketverlust ist mit hilfe des SNR abgeschützt. Im Artikel sind beschrieben die Ergebnisse aus der Simulation fur die Veranschlagung SNR im Fall des konkretes Wahl bestimmenes bazis Funktion.