USEFULNESS OF SPEECH CODING IN VOICE BANKING

M. Satya Sai Ram m_satyasairam@yahoo.co.in
Chalapathi Institute of Technology
Guntur, AP, INDIA

P. Siddaiah siddaiah_p@yahoo.com
K.L.College of Engineering
Guntur, AP, INDIA

M. Madhavi Latha mlmakkena@yahoo.com
J.N.T.U College of Engineering
Hyderabad, A.P, INDIA

Abstract

Voice banking is an excellent telephone banking service by which a user can access his account for any service at any time of a day, in a year. The speech techniques involved in voice banking are speech coding and speech recognition. This paper investigates the performance of a speech recognizer for a coded output at 20 bits/frame obtained by using various vector quantization techniques namely Split Vector Quantization, Multistage Vector Quantization, Split-Multistage Vector Quantization, Switched Split Vector Quantization using hard decision scheme, Switched Multistage Vector Quantization using soft decision scheme and Multi Switched Split Vector Quantization using hard decision scheme techniques. The speech recognition technique used for recognition of the coded speech signal is the Hidden Markov Model technique and the speech enhancement technique used for enhancing the coded speech signal is the Spectral Subtraction technique. The performance of vector quantization is measured in terms of spectral distortion in decibels, computational complexity in Kflops/frame, and memory requirements in floats. The performance of the speech recognizer for coded outputs at 20 bits/frame has been examined and it is found that the speech recognizer has better percentage probability of recognition for the coded output obtained using Multi Switched Split Vector Quantization using hard decision scheme. It is also found that the probability of recognition for various coding techniques has been varied from 80% to 100%.

Keywords: Voice banking, Product Code Vector Quantizers, Linear Predictive Coefficients, Line Spectral Frequencies.

1. INTRODUCTION

Among number of applications for Speech coding, one of its main application is its use in voice banking. Voice banking is an excellent telephone banking service, which makes the user to be in touch with his account information at any time of the day. Speech coding plays an important role in voice banking, which involves the recognition of the coded outputs from a telephone line for performing a particular banking operation. The speech recognizer at the banking end tries to understand the compressed spoken words in some way or the other and will act thereafter.

This paper takes the advantage of voice banking application and examined the performance of a speech recognizer for the coded outputs obtained by using various vector quantization techniques at 20bits/frame. The vector quantization techniques used in this work are the Split Vector Quantization (SVQ) [1-2], Multistage Vector Quantization (MSVQ) [3], Split-Multistage Vector Quantization (S-MSVQ) [4], Switched Split Vector Quantization (SSVQ) using hard decision scheme[3], Switched Multistage Vector Quantization (SWMSVQ) using soft decision scheme [5] and Multi Switched Split Vector Quantization (MSSVQ) using hard decision scheme [6-7].

Voice Banking is very useful in daily life, an excellent telephone banking service that makes the user to have an access to his account information, and other banking services 24 hours a day, 7 days a week and 365 days a year, just by making a simple phone call. With voice banking the user can have an access to the following services:

1) The user can check for balances on all his current, savings, and loan accounts.
2) The user can know the history of his account like the recent deposits, debits, check payments and interests.
3) The user can make funds transfer between two accounts and can make loan payments.
4) The user can check for interest rates on deposits and on various types of loans.
5) The user can stop the payments to be made.
6) The user can make a report about the lost or stolen card.
7) The user can have information about the bank branches, ATM locations and working hours.
8) The user can speak to the banks customer service center at any time of the day.

various other services can also be provided to the customers depending on the bank.

The speech techniques involved in voice banking are the speech coding, speech enhancement and speech recognition. Speech coding is used to compress the words spoken by the user before transmitting them over a telephone line. The speech enhancement techniques are used to remove the noise content before and after compression prior to applying them to a speech recognizer at the banking end. The speech recognition technique is used to recognize the compressed words spoken by the user. The speech parameters used for coding are the Line Spectral Frequencies (LSF) [8-9], so as to ensure filter stability after quantization. The parameters that can be used for speech recognition are the Linear Predictive Coefficients (LPC) and Mel-Frequency Cepstral Coefficients (MFCCs). In this work the parameters used for recognition are the Linear Predictive Coefficients. To improve the performance of recognition Energy, Delta and Acceleration coefficients must be used, but in this paper they are not used because if they are used, the generation of codebooks becomes a difficult task as the number of samples per vector increases.

The speech recognition technique used for the recognition of the coded outputs is the Hidden Markov Model (HMM) technique [10-11]. HMM is a collection of various statistical modeling techniques, in which the transition probability matrix is estimated with the help of the Baum Welch algorithm [11], the emission matrix is generated by using the K-means clustering algorithm and is estimated using the Baum Welch algorithm. The Viterbi algorithm can also be used for the estimation of the transition and emission matrices. For a given sequence the most likely sequence path is estimated using the Viterbi algorithm [11], from which the probability of a particular sequence is estimated using the forward algorithm or the backward algorithm. The stages involved in voice banking are: 1) Top level dialogue structure 2) Login mode 3) Account Balance Sub-Dialogue 4) Funds Transfer Sub-Dialogue [12].

2. TOP LEVEL DIALOGUE STRUCTURE

The Flow chart of a Top level dialogue structure is shown in Figure 1. The Top level dialogue structure consists of the following steps:

1) In the first step the voice banking system asks for the account number and PIN. When the user enters a valid account number and PIN, the voice banking system says that the 'Login is successful' and allows the user to enter the voice banking system or else it says the 'Login is unsuccessful'.
2) In the second step the voice banking system asks the user to select the required operation. If the user asks for the account balance, the account balance sub dialogue will be processed, if the user asks for funds transfer the funds transfer sub dialogue will be processed.
3) After performing the required banking operation the voice banking system asks the user 'Would you like to do any further operation?' If the user says 'Yes' the operation will go to step 2, if the user says 'No' it will say thank you for calling to voice banking and the process will be terminated.

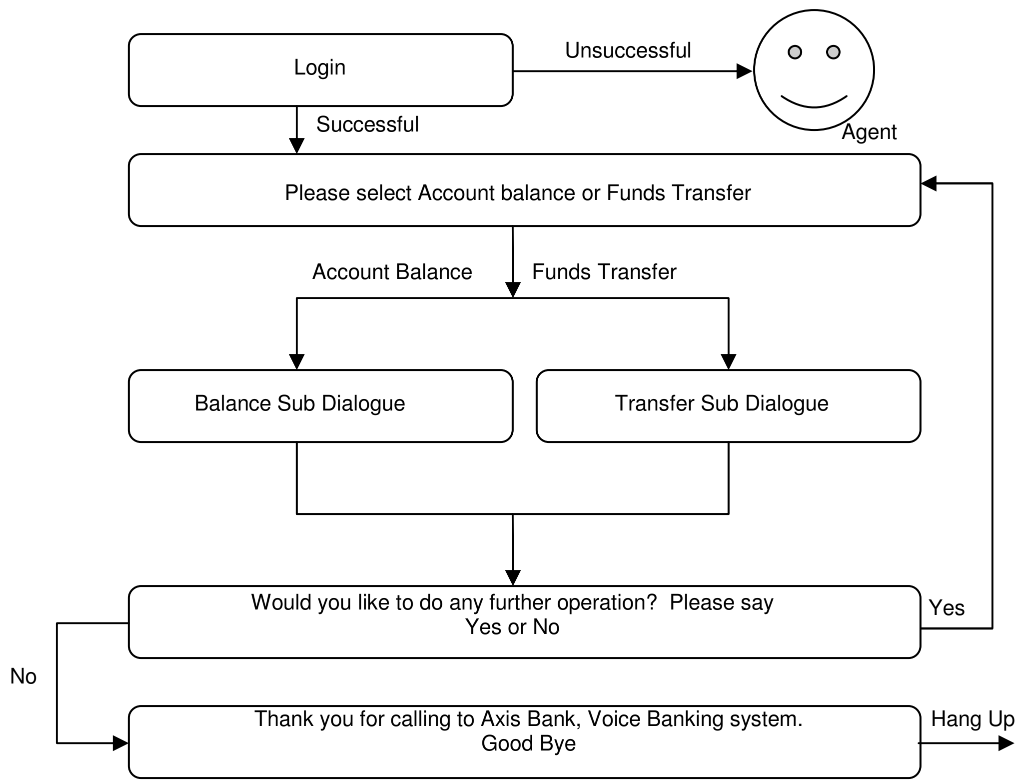

Figure 1: Top Level Dialogue Structure

3. LOGIN MODE

The Flow chart of the Login Mode in voice banking is shown in the Figure 3. The Login Mode consists of the following steps:

1. When a user dials for voice banking, a Welcome dialog appears and the voice banking system asks the customer to tell the account number.
2. When the customer tells the account number the voice banking system checks the account number for validation.
3. If the account number is valid, the voice banking system asks the customer to tell the PIN number, else it says that the account number is invalid and asks to try again. The account number given by the customer must be recognized by the voice banking system in three attempts otherwise the account will be blocked by the banking system.

4. When the account number and PIN are valid the voice banking system allows the customer to perform the required transaction by saying that the 'Login is successful'.

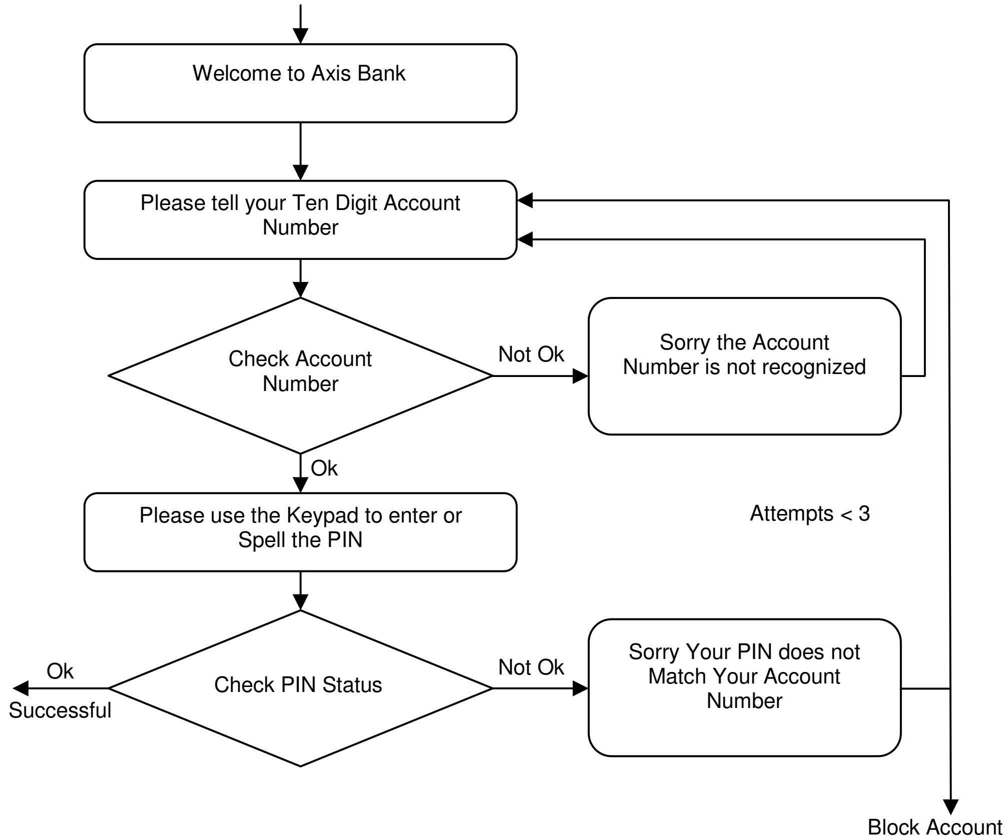

Figure 3: Login Mode

4. ACCOUNT BALANCE SUB-DIALOGUE

The Flow chart of the Account Balance sub-dialogue in voice banking is shown in the Figure 2.

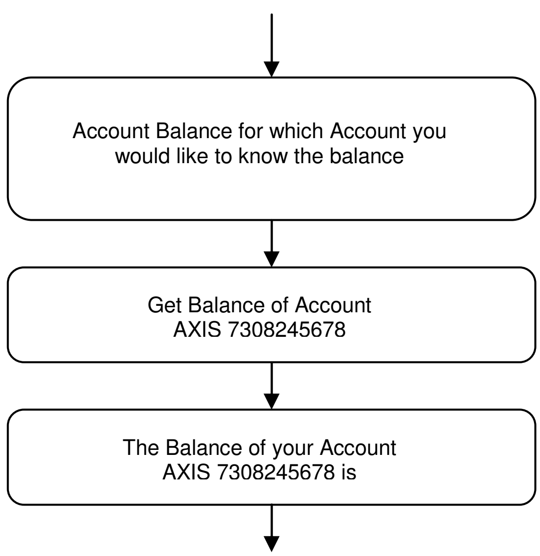

Figure 2: Account Balance sub-dialogue

The Account Balance sub-dialogue consists of the following steps:

1. When the user asks for account balance, the voice banking system asks for the type of account for which the user likes to know the balance.
2. When the user tells the type of account and account number, and if the account number is valid then the voice banking system will tell the balance of the account.

5. FUNDS TRANSFER SUB-DIALOGUE

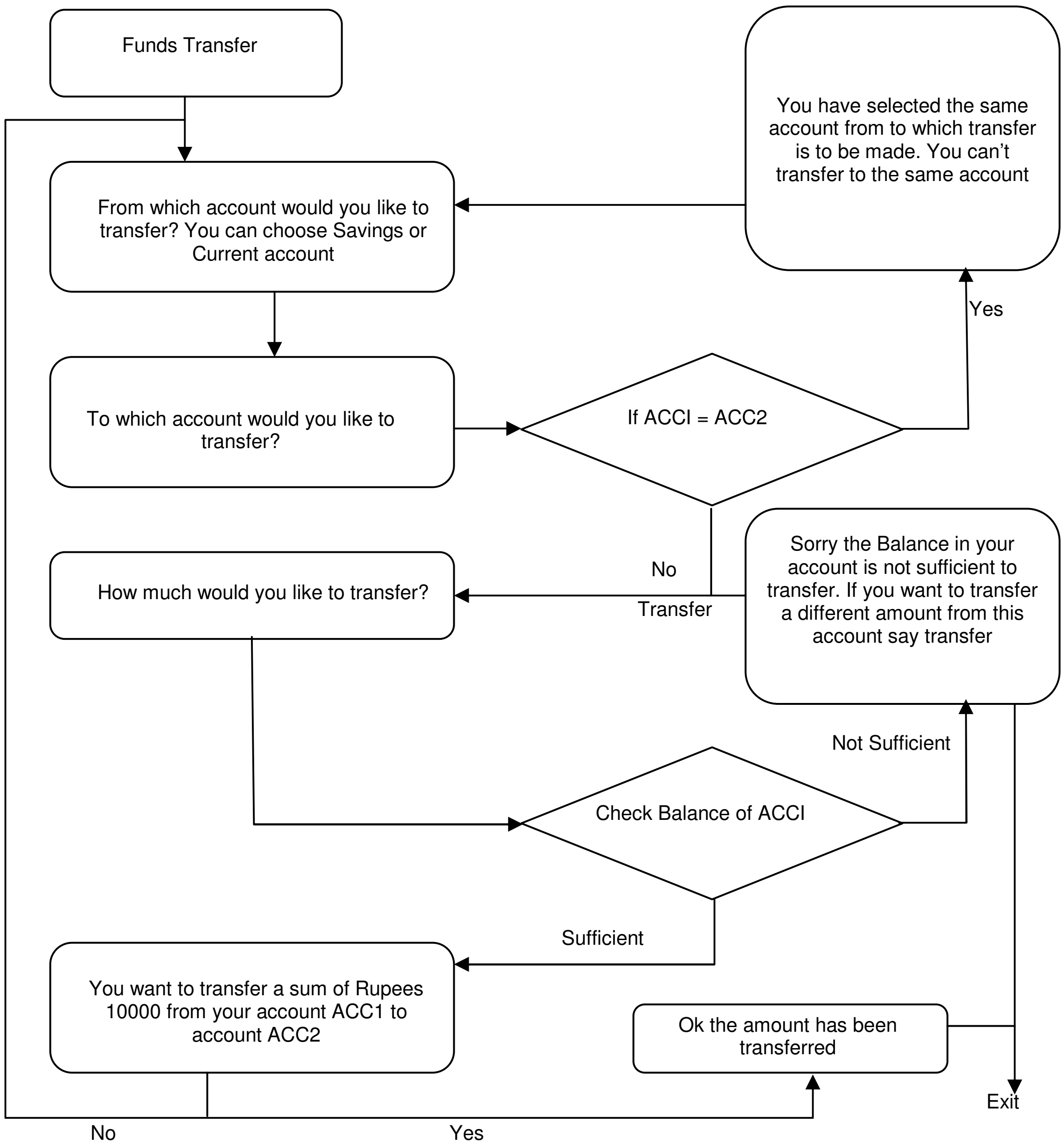

Figure 4: Funds Transfer sub-dialogue

The Flow chart of the Funds Transfer sub-dialogue in voice banking is shown in Figure 4. It consists of the following steps:

1. In the first step of funds transfer the voice banking system will ask the user from which account you would like to transfer, either savings or current account.
2. In the second step it asks for the account to which transfer is to be made.
3. In the third step when the user tells the account number, the voice banking system checks the account number to which transfer is to be made.
4. If both the account numbers to which transfer is to be made are same then the voice banking system says that you cannot transfer to the same account and the process will return back to step1.
5. If the two account numbers are not identical then the voice banking system asks 'How much would you like to transfer?'.
6. In the fourth step when the user speaks for the amount to be transferred, the voice banking system checks the balance of the account from which transfer is to be made.
7. If the balance is not sufficient, the voice banking system says 'There is insufficient balance in the account' and asks to enter a different amount.
8. If the balance in the account is sufficient, the banking system says 'Do you want to transfer the amount from ACC1 to ACC2 ?' (Account1 to Account2).
9. When the user says 'Yes' the transfer will be made and if the user says 'No' it will return to step1.
10. When the transfer is made, the banking system says 'The transfer has been made' and the process will be terminated.

The compressed speech output must be of good quality, otherwise the speech recognizer may fail to recognize the spoken words, if it fails to recognize in three attempts the account will be blocked. So an efficient coding technique is required to compress the spoken words. In this work the performance of the speech recognizer has been observed for various coded outputs, obtained by using various vector quantization techniques at 20 bits/frame. The quantization techniques used for coding are SVQ, MSVQ, S-MSVQ, SSVQ using hard decision scheme, SWMSVQ using soft decision scheme, and MSSVQ using hard decision scheme techniques. The speech enhancement technique used is the 'Spectral subtraction' technique, it is used to remove the noise in the coded speech signal after transmission and the speech recognition technique used is the 'Hidden Markov Model technique'. The steps involved in speech coding, enhancement and recognition to obtain a good quality of speech for the voice bank recognizer are:

1) In the first step, the silence part of the speech signal must be removed by using the voice activation and detection technique.
2) In the second step, the speech signal must be coded using a vector quantization technique.
3) In the third step, the coded output with added channel noise must be enhanced by using an enhancement technique.
4) In the fourth step, the enhanced speech signal must be given as an input to the voice bank recognizer for recognizing.
5) Finally the probability of recognition can be computed as a measure of the recognition accuracy.

From results it has been observed that for the coded outputs obtained by using various vector quantization techniques and the recognition accuracy has been varied from 80% to 100%.

6. MULTI SWITCHED SPLIT VECTOR QUANTIZATION

The block diagram of a p x m x sp Multi Switched Split Vector Quantizer is shown in Figure 5. In Figure 5, p corresponds to the number of stages, m corresponds to the number of switches, and sp corresponds to the number of splits. Each input vector's' to be quantized, is applied to the first stage of a Multi Switched Split Vector Quantizer to obtain the quantized version of the input vector 's' given by $\hat{s}_1 = Q[s]$. The error vector at the first stage $e_1 = s - \hat{s}_1$ will be computed and is quantized using the Switched Split Vector Quantizer at the second stage to obtain the quantized version of the error vector $\hat{e}_1 = Q[e_1]$. This process can be continued for the required number of stages. Finally the decoder takes the indices, I_i, from each stage and adds the quantized vectors at each stage to obtain the quantized version of the input vector s given by $\hat{S} = Q[S] + Q[e_1] + ...$, where $Q[S]$ is the quantized version of the input vector at the first stage, $Q[e_1]$ is the quantized version of the error vector at the second stage and so on [6-7].

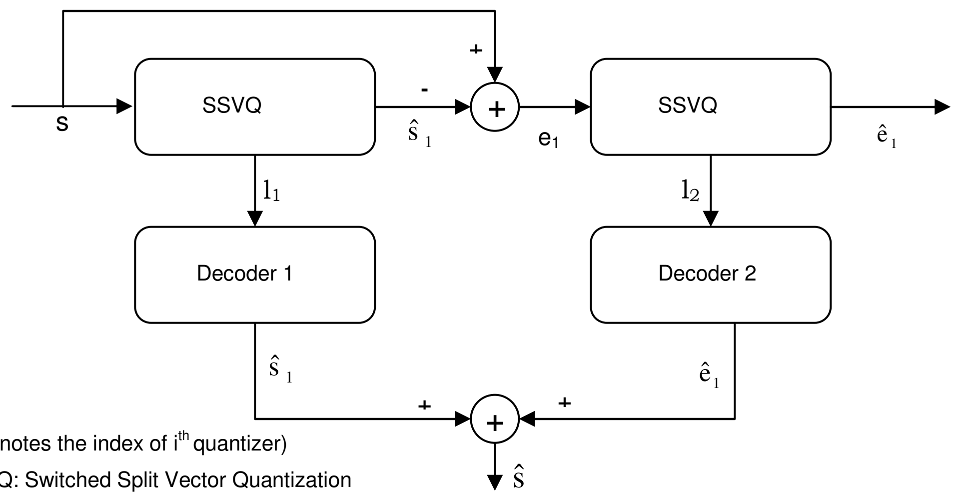

Figure 5: Block Diagram of MSSVQ

7. RESULTS

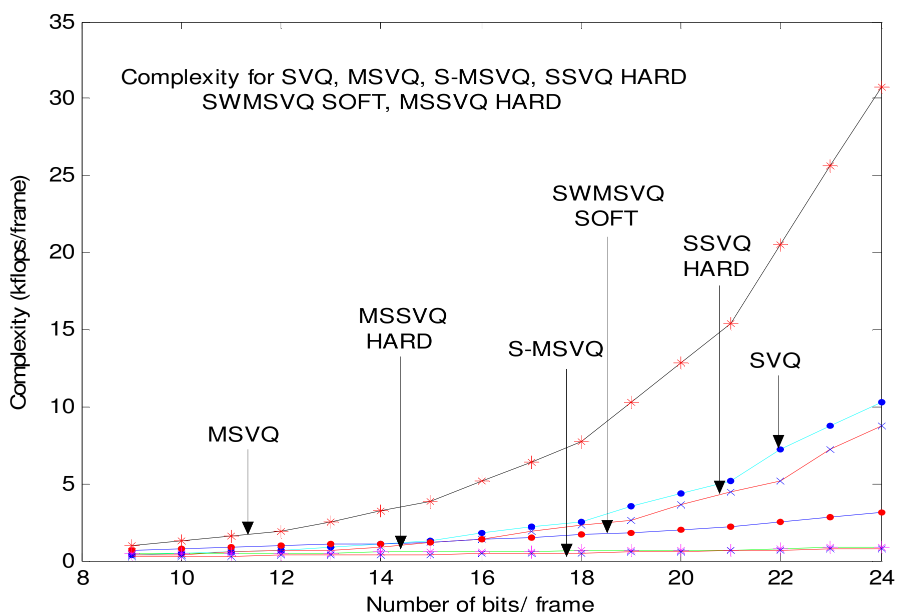

Figure 6: Complexity for SVQ, MSVQ, S-MSVQ, SSVQ hard, SWMSVQ soft, and MSSVQ hard

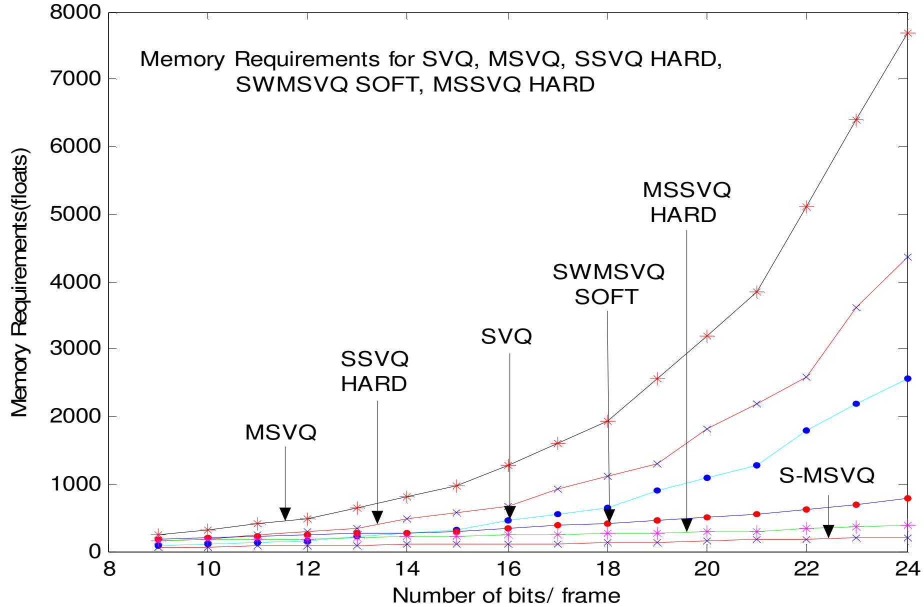

Figure 7: Memory requirements for SVQ, MSVQ, S-MSVQ, SSVQ hard, SWMSVQ soft, and MSSVQ hard

Table 1: Spectral distortion, Complexity, and Memory requirements for 3-part Split Vector Quantization

Bits / frame	SD (dB)	Percentage of outliers		Complexity (kflops / frame)	ROM (floats)
		2-4 dB	>4 dB		
24(8+8+8)	1.45	0.43	0	10.237	2560
23(7+8+8)	1.67	0.94	0	8.701	2176
22(7+7+8)	1.70	0.78	0.1	7.165	1792
21(7+7+7)	1.83	2.46	0.2	5.117	1280
20(6+7+7)	1.81	1.6	0.4	4.349	1088

Table 2: Spectral distortion, Complexity, and Memory requirements for 3-stage Multistage Vector Quantization

Bits / frame	SD (dB)	Percentage of outliers		Complexity (kflops / frame)	ROM (floats)
		2-4 dB	>4 dB		
24(8+8+8)	0.984	1.38	0	30.717	7680
23(8+8+7)	1.238	1.2	0.1	25.597	6400
22(8+7+7)	1.345	0.85	0.13	20.477	5120
21(7+7+7)	1.4	1.08	0.3	15.357	3840
20(7+7+6)	1.41	1.1	0.4	12.797	3200

Table 3: Spectral distortion, Complexity, and Memory requirements for 3-part, 3-stage Split-Multistage Vector Quantization

Bits / frame	SD (dB)	Percentage of outliers		Complexity (kflops / frame)	ROM (floats)
		2-4 dB	>4 dB		
24(8+8+8)	0.0345	0	0	0.807	204
23(8+8+7)	0.0385	0	0	0.759	192
22(8+7+7)	0.0378	0.1	0	0.711	180
21(7+7+7)	0.0382	0.2	0	0.663	168
20(7+7+6)	0.0389	0.3	0	0.599	0.152

Table 4: Spectral distortion, Complexity, and Memory requirements for 2- switch 3-part Switched Split Vector Quantization using hard decision scheme

Bits / frame	SD (dB)	Percentage of outliers		Complexity (kflops / frame)	ROM (floats)
		2-4 dB	>4 dB		
24(1+7+8+8)	0.957	1.06	0	8.78	4372
23(1+7+7+8)	1.113	1.29	0.14	7.244	3604
22(1+7+7+7)	1.119	0.52	1.3	5.196	2580
21(1+6+7+7)	1.127	1.3	0.56	4.428	2196
20(1+6+6+7)	1.09	1.3	0.63	3.66	1812

Table 5: Spectral distortion, Complexity, and Memory requirements for 3-stage 2-switch Switched Multistage Vector Quantization using soft decision scheme

Bits / frame	SD (dB)	Percentage of outliers		Complexity (kflops / frame)	ROM (floats)
		2-4 dB	>4 dB		
24(1+7+1+7+1+7)	0.91	0.56	0.81	3.111	780
23(1+7+1+7+1+6)	0.87	1.05	0.31	2.791	700
22(1+7+1+6+1+6)	1.1	1.45	0.63	2.471	620
21(1+6+1+6+1+6)	1.18	0.6	1.89	2.151	540
20(1+6+1+6+1+5)	1.15	0.69	1.83	1.99	500

Table 6: Spectral distortion, Complexity, and Memory requirements for 2-switch 3-part 3-stage Multi Switched Split Vector Quantization using hard decision scheme

Bits / frame	SD (dB)	Percentage of outliers		Complexity (kflops / frame)	ROM (floats)
		2-4 dB	>4 dB		
24(1+7+1+7+1+7)	0.0322	0	0	0.9	396
23(1+7+1+7+1+6)	0.0381	0	0	0.836	364
22(1+7+1+6+1+6)	0.0373	0	0	0.772	332
21(1+6+1+6+1+6)	0.0377	0	0	0.708	300
20(1+6+1+6+1+5)	0.0376	0	0	0.684	288

Table 7: Probability of recognizing a word ONE at 20 bits / frame by using SVQ, MSVQ, S-MSVQ, SSVQ using hard decision scheme, SWMSVQ using soft decision scheme and MSSVQ using hard decision scheme

NAME	PROBABILITY OF RECOGNITION					
	SVQ	MSVQ	S-MSVQ	SSVQ	SWMSVQ	MSSVQ
ZERO	-16.4401	-14.3510	-11.0312	-17.3314	-15.639	-10.3625
ONE	-19.6432	-16.8930	-19.3211	-20.8234	-21.3025	-19.1769
TWO	-19.6941	-16.9001	-14.6350	-15.0031	-15.918	-14.220
THREE	-16.3331	-13.1160	-12.0561	-17.5689	-17.493	-11.1311
FOUR	-15.9137	-15.6603	-17.3426	-17.7058	-16.749	-16.9621
FIVE	-17.0071	-17.0531	-17.1076	-19.1743	-17.069	-16.5635
SIX	-18.0012	-14.5134	-14.1384	-18.9013	-18.938	-13.3014
SEVEN	-15.9981	-13.0007	-13.4420	-15.1007	-19.687	-11.6120
EIGHT	-14.8371	-14.5563	-13.8603	-15.2330	-11.806	-12.1421
NINE	-19.6801	-17.1281	-20.1107	-21.1163	-20.031	-19.106
TEN	-18.9561	-15.0081	-18.3214	-18.4418	-13.671	-17.5038
YES	-16.8341	-14.3842	-17.2014	-17.3341	-20.398	-16.6141
NO	-20.3259	-14.9802	-18.6345	-21.9276	-16.943	-18.0170
SUCESSFULL	-13.9253	-10.3310	-11.1010	-14.0140	-10.621	-9.6210
UNSUCESSFULL	-13.1109	-10.5676	-11.0048	-15.0121	-11.006	-9.7128
% RECOGNITION	80.00%	80.00%	93.33%	86.66%	100%	100%

The computational complexity, and memory requirements of MSSVQ using hard decision scheme is less when compared to all the product code vector quantization techniques except for S-MSVQ which can be observed from Tables 1-6 and from Figure's 6 and 7 . Table 7 shows the probability of recognizing a particular coded word 'ONE' obtained by using SVQ, MSVQ, S-MSVQ, SSVQ using hard decision scheme, SWMSVQ using soft decision scheme and MSSVQ using hard decision scheme at 20 bits/ frame. The reason for choosing 20 bits/frame is that with MSSVQ using hard decision scheme the transparency in quantization has been achieved at 20 bits/frame, so 20 bits/frame is taken as the reference. From Table 7 it can be observed that for the utterance 'ONE' as an input the probability of recognition is better for SWMSVQ using soft decision scheme and MSSVQ using hard decision scheme and it is 100%. But it is advantageous to use MSSVQ using hard decision scheme for voice banking application as its spectral distortion is less when compared to all the product code vector quantization techniques which can be observed from tables 1 to 6. So it is proved that MSSVQ using hard decision scheme can be better used in voice banking applications for coding of the speech signals.

8. CONCLUSION

The Speech recognizer using HMM performs well for the coded output obtained by using MSSVQ using hard decision scheme at 20 bits / frame, as it has less spectral distortion. MSSVQ can have better marketability as it has less computational complexity and memory requirements. From results it is observed that the probability of recognition has been varied from 80% to 100% for various vector quantization techniques and for Switched Multistage Vector Quantization using soft decision scheme and Multi Switched Split Vector Quantization using hard decision scheme the probability of recognition is 100%. But it is proved that Multi Switched Split Vector Quantization technique using hard decision scheme is superior in terms of spectral distortion, computational complexity and memory requirements when compared to other product code vector quantization techniques. So Multi Switched Split Vector Quantization using hard decision scheme is proved to be better and is having performance closer to Split-Multistage Vector Quantization technique. So Multi Switched Split Vector Quantization using hard decision scheme is proved to be the better coding technique for voice banking application. The performance can be further improved by increasing the number of training vectors, bits used for codebook generation, the number of states of an utterance, by using an efficient algorithm for the generation of emission matrix that takes into account the entire training set unless the K-means clustering that randomly picks vectors from the training set for the generation of an emission matrix, and by using a software having greater degree of precision. With ‘Matlab’ the generation of an emission matrix with k-means is difficult when the number of states is more for a particular utterance.

9. REFERENCES

1. Paliwal K.K, Atal, B.S. “Efficient vector quantization of LPC Parameters at 24 bits/frame”. in: Proc.IEEE Int. Conf. Acoust, Speech, Signal Processing, (1): 661-664,1991.
2. Paliwal. K.K, Atal, B.S. “Efficient vector quantization of LPC Parameters at 24 bits/frame”. IEEE Trans. Speech Audio Process, 1(1): 3-14, 1993.
3. Stephen, So., Paliwal, K. K. “Efficient product code vector quantization using switched split vector quantizer”. Digital Signal Processing journal, Elsevier, 17(1): 138-171,2007.
4. Woo-Jin Han, Eun-Kyoung Kim, Yung-Hwan Oh. “Multi codebook Split Vector Quantization of LSF parameters”. IEEE Trans. Signal Processing, 9: 418-421, 2002.
5. K.Srinivas, M.Madhavi Latha, P.Siddaiah, M.satya Sai Ram.(July 2008a). Switched Multi stage vector quantization using soft decision scheme. Proceedings of the 2008 International conference on Image processing, computer vision and pattern recognition (IPCV 2008), WORLD COMP2008, 2: 587-591, 2008.
6. M.Satya Sai Ram, P.Siddaiah, M.MadhaviLatha. “Multi Switched Split Vector Quantization of Narrow Band Speech Signals”, Proceedings World Academy of Science, Engineering and Technology, WASET, 27: 236-239,2008.
7. M.Satya Sai Ram, P.Siddaiah, M.MadhaviLatha. “Multi Switched Split Vector Quantizer”, International Journal of Computer, Information, and Systems science, and Engineering, IJCISSE, WASET, 2(1): 1- 6, 2008.
8. Kleijn, W.B. Tom Backstrom, Paavo Alku.”On Line Spectral Frequencies”. IEEE Signal Processing Letters, 10(3): 75-77, 2003.
9. Soong. F, Juang. B. “Line spectrum pair (LSP) and speech data compression”, IEEE Conference. On Acoustics, Speech Signal Processing, 9(1): 37-40, 1984.
10. Rabiner Lawrence, Juang Bing-Hwang. “Fundamentals of speech Recognition”. Prentice Hall, New Jersey, ISBN 0-13-015157-2, 1993.
11. Rabiner L.R. “A tutorial on Hidden Markov Models and selected applications in speech recognition”. Poceedings of the IEEE, 77(2): 257-286, 1989.
12. Robert Dale. “Dialog Design Specification: A Voice Banking Application,” Macquarie University - Sydney, 0.1: 1-22, 2002.