음성 전보 시스템 개발의 연구 동향

민 소 연, 이 수 민, 이 순 규, 배 명 진
숭실대학교 정보통신공학과
156-743 서울시 동작구 상도동 1-1
Phone:820-0902 Fax:826-0461

On Reseach trends of Voice Telegram System(VTS)

SoYeon Min, SoonMin Lee, SunGyu Lee, MyungJin Bae
Dept. Information and Telecommunication Engr., Soongsil University
1-1 Sangdo-5Dong, Dongjak-Ku, Seoul 156-743, KOREA
symin@assp.soongsil.ac.kr
mjbae@saint.soongsil.ac.kr

Abstract

Telegram is an indispensible information & telecommunication system to our daily life. **VTS**(**V**oice **T**elegram **S**ystem) under intensive research is intended to enhance exchanging capability of information & telecommunication by adding voice media to existing telegram system. Overall configuration and necessary core technologies of the system were investigated for its development. Among those many technologies in need, the technology of compressing and recording data is most critical to the development of cheap hardware. This is so called vocoder algorithm and is the core technology of voice information system. So, here, vocoder algorithm now being studied will be introduced.

1. 서 론

우리가 연구-개발하고 있는 VTS는 기존의 전보전송체계에 음성매체를 포함시킴으로서 전보에 의한 정보통신 교환능력을 향상시키는 것이다. 전보 통신분야에서 문자 정보서비스에 목소리 정보 서비스를 함께 제공하면 메시지의 정확한 전달이 이루어지고, 발신자의 주변환경이 기록될 수 있으므로 전보통신에 멀티미디어 기술이 접목될 수 있는 길을 여는 것이 된다.

그림 1-1은 음성 전보 시스템의 구성도이다.
먼저 사용자가 인터넷이나 전화 또는 직접 방문하여, 전보를 신청하여 전하고자 하는 음성 메시지를 녹음 하면 음성 압축을 수행하여 데이터 량을 줄이게 된다.

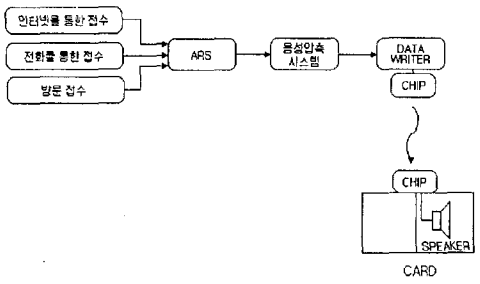

그림 1-1. 음성 전보 시스템 구성도

수신자가 카드를 펼치면 기록된 음성 메시지는 칩속에 프로그램된 음성 복호화 알고리즘에 의해 다시 복원되어 보낸 사람의 목소리로 메시지를 전달하게 된다.

그림 1-2는 음성 전보 시스템의 사용 예를 보인 것이다. 음성 전보 시스템의 주요 핵심기술 내용은 크게 음성 압축 및 복원을 수행하는 보코더 알고리즘 기술과 이러한 보코더 알고리즘을 저렴한 하드웨어로 구현하는 기술이다. 하지만 이러한 큰 기술 이외에도 음성 전보 시스템을 사용자가 쉽게 사용할 수 있게 하는 공중망 접속기술 및 여러 중요한 기술이 필요하게 된다.

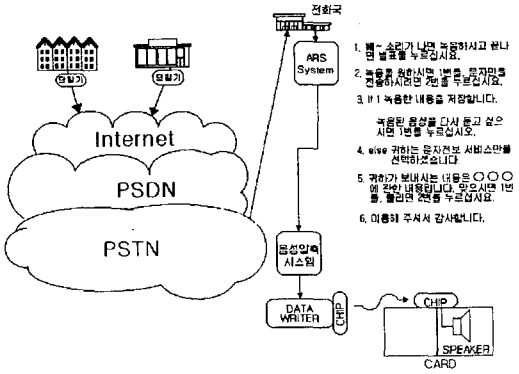

그림 1-2. 보이스 전보 시스템 사용예

다음은 VTS 소프트웨어의 핵심기술 개념도 이다.

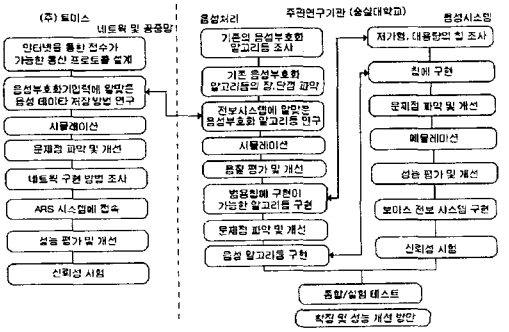

그림 1-3. 핵심기술 소프트웨어 개념도

2. 음성 전보 시스템의 CELP형 보코더

ITU-T에서 표준화한 압축률이 높고 음질이 우수한 CELP 계열의 특성을 간단히 살펴본다[1].

2-1. 6.3/5.3Kbps MP-MLQ/ACELP(G.723)

G.723의 기본적인 구조는 기존 CELP와 같다. 여기신호의 구성과 비트 할당에 있어서 약간 다르며 그 특징을 요약하면 다음과 같다.

1. 8kHz 샘플링, 16 비트 linear PCM 입력 신호.
2. 30ms 프레임 길이 + 7.5ms look ahead = 37.5ms codec delay.
3. 30ms LPC 분석, 15ms마다 피치 검색 및 갱신, 7.5ms마다 코드북검색
4. 5.3kbps와 6.3kbps의 두 가지 전송률 스위칭 : 5.3kbps ACELP(Algebraic CELP), 6.3kbps MP-MLQ (Multipulse Maximum Likelyhood Quantization).
5. Harmonic noise shaping.
6. 선형 예측 분석-합성(analysis-by-synthesis).

2-2. 8kbps CS-ACELP(G.729)

1995년에 ITU-T에서는 기존의 표준안들인 32kbps ADPCM(G.726)과 16kbps LD-CELP(G.728)와 거의 유사한 음질을 갖는 저 전송률 음성 부호화 표준안 8kbps CS-ACELP(Conjugate Structure Algebraic CELP: G.729)을 결정하였다.

▶ G.729의 주요한 특징을 보면 다음과 같다:

1. 8㎑ 샘플링, 1/2 다운 스케일링.
2. 15㎳의 지연 시간 = 10㎳ frame + 5㎳ look ahead.
3. 채널 잡음 하에서 G.726에 상응하는 음질.
4. 개회로 + 폐회로 피치 검색, 델타 피치 부호화.
5. 다단 + 분리 VQ를 이용한 LSP 부호화.
6. MA(Moving Average) 이득 예측.
7. 피치 프리필터링, 적응 포스트필터링.

2-3. 4Kbps 음성 부호화 표준안

저 전송률 비디오 폰과 같이 멀티미디어 응용 시스템과 함께 늘어나는 이동통신의 수요 8kbps 이하의 낮은 전송률을 갖는 부호화 알고리즘을 필요로 하게 하였다. <표 2-1>은 4Kbps 음성 부호화기의 요구 사항 이다.

<표. 2-1 4Kbps 음성 부호화기의 요구사항>

Parameter	Requirement(s)
Speech quality in error free condition at nominal input level of -26.15db with respect to the OVL point(-20dbm0)	Not worse than that of ITU-T Rec. G.726 at 32 kbit/s at nominal input level of -26.15 dB with respect to the OVL point (-20dBm0)
Speech quality dependency on the input signal level between -36.15dB and -16.15dB with respect to the overload point	Not worse than that of ITU-T Rec. G.726 at 32kbit/s
Quality dependency on speakers	Not worse than that of ITU-T Rec. G.726 at 32kbit/s

3. 음성 전보 시스템의 보코더 알고리즘에 관한 연구

3.1 Skipping 기법을 이용한 피치검색 단축

G.723.1은 저 전송률 환경에서 고음질을 제공하여 주고 있으나 CELP형 부호화기가 갖는 합성에 의한 분석(analysis by synthesis) 방식의 구조로 인해 많은 처리 시간과 계산량을 요구하게 된다. 이 연구에서는 G.723.1에 대해 skipping 기법을 이용하여 피치 검색 과정의 계산량을 줄여 부호화기의 전체 처리 시간을 감소시키는 방법을 제안하였다[2]. 예측 피치를 찾기 위한 개회로 피치 예측(open loop pitch estimation) 과정에서 계산량을 줄이기 위해 skipping 기법을 사용 하였다. 피치 예측 과정시 상관관계 파형은 양과 음의 파형이 교대로 나타나는 특징을 가지고 있기 때문에 계산시 음의 파형을 생략하는 방법을 사용하였다[3].

실제 음성시료에 대해 제안한 피치 검색법을 적용하였을 때 부호화시 평균 처리시간은 약 10%정도 감소하였으며 기존 G.723.1과 제안한 방법을 적용한 G.723.1의 음질 비교를 위하여 MOS 평가를 했을 때 기존의 방법이 평균 3.76인데 비해 제안한 방법의 평균 MOS는 3.73으로 주관적인 음질 저하는 거의 나타나지 않았다[4].

3.2 다중 레벨 진폭 대수 코드북을 이용한 음성 부호화기에 관한 연구

기존의 대수코드북 구조를 보면 몇가지의 문제점을 발견할 수 있다[5][6].

1. 멀티펄스에 대해 가능성이 있는 일부 펄스 위치만을 검색하는 부분적인 하이퍼큐브 방식을 사용한다.
2. 각 트랙에서 단지 하나의 펄스만을 검색하므로 동일 트랙상에 다른 위치에 있는 두 펄스가 최적의 값을 갖더라도 준최적을 이루는 펄스를 검색한다.
3. 발견된 모든 펄스가 일정한 진폭을 가지고 있다고 가정하고 ±1이라는 규준화된 진폭만을 검색한다.

따라서 본 연구에서는 이를 개선하기 위하여 다음과 같은 코드북을 구성하였다. 검색방법은 DFS검색을 사용한다.

<표. 3-1 개선을 위해 사용된 코드북>

트 랙	부 호	위 치
T_0	±1,±0.5	0,8,16,24,32,40,48,56
T_1	±1,±0.5	2,10,18,26,34,42,50,58
T_2	±1	4,12,20,28,36,44,52,(60) 6,14,22,30,38,46,54,(62)

첫 번째와 두 번째의 문제 해결을 위해, 트랙 1, 2에서는 한 펄스를 각 트랙에서 검색하여 발견하며 트랙 3에서는 두개의 펄스를 검색한다. 왜냐하면 트랙 3에서 다른 위치에 있는 펄스가 최적으로 판단되었을 경우 검색 불가능한 것을 방지하기 위함이다. 또한 동일 위치에 두 펄스가 최적이라 판단되었을 경우 진폭부호는 두 펄스의 합인 진폭 ±2를 전송한다.

세 번째 문제를 해결하기 위하여 다음의 방식을 제안하였다. 기존대수적 코드북의 단점은 펄스진폭비트를 ±1의 부호 성분만 나타내어 다양한 펄스 진폭 변화를 나타낼 수 없었다. 또한 각 펄스마다 N의 진폭변화를 설정할때 트랙이 L개 인 경우 Σ펄스위치+(N×L)만큼의 비트가 필요하다. 즉, 코드북 크기의 급격한 증가로 검색 시 많은 계산 량이 필요로 된다.

이를 개선하기 위해서 먼저 부프레임 내에서의 펄스진폭을 미리 설정한다. 식(3.1)로 진폭을 결정하기 위한 목적 진폭신호를 만든다. 식(3.2)는 펄스 설정을 위해 목적벡터와 피치성분을 제거한 원음성 그리고 LPC성분을 제거한 원음성과의 적절한 혼합 비율로 구성되어있다. 여기서 $\upsilon_1, \upsilon_2, \upsilon_3$는 0과 1사이의 값이어야 하며 본 연구에서는 $\upsilon_1=\upsilon_2=0.5$, $\upsilon_3=0.1$을 사용하였다. 결과로 추출된 펄스 진폭은 부프레임 구간 내에서 식(3.2)를 이용하여 정규화한다.

$$Y = \upsilon_1 \frac{D}{||D||} + \upsilon_2 \frac{P}{||P||} + \upsilon_3 \frac{S}{||S||} \qquad (3.1)$$

$$\Delta = Q\left(\frac{V_P}{\max_k |V_n|} \right) \qquad (3.2)$$

max $|V_n|$는 영이 아닌 진폭펄스의 최대치를 나타내는 정규화 인자이고 Q(·)는 양자화 함수이다[3][7].

본 연구에서는 ±1/±0.5 크기의 진폭을 사용하므로 트랙 2의 경우 선택된 진폭에서 부호 비트만을 사용한다. 트랙 1,2에서는 Δ함수가 미리 설정된 문턱값에 따라 ±1/±0.5의 진폭을 결정한다. 검색 과정은 기존의 Depth-First Search법과 동일하다. 단, 행렬 $\Phi(i,j)$를 계산하는데 제안된 알고리즘으로 미리 계산되어진 진폭을 식(3.3)과 같이 결합시킨다. 그 결과로 진폭이 q개의 경우의 수를 가지더라도 진폭을 검색하기 위한 계산량의 증가는 발생하지 않는다.

$$\Phi(i,j) = S_i S_j \Phi(i,j) \qquad (3.3)$$

시뮬레이션 결과, 로그영역비를 이용한 무성음에서 유성음으로의 천이검출과 제안된 대수 코드북을 이용하여

펄스진폭의 다양성을 개선 하였으며, 위치 선택에 있어서의 제한을 비트증가없이 완화시켰다. 대수 코드북에서는 코드북의 구조를 개선하여 2개의 펄스를 한 트랙에서 검출하여 대수코드북이 가지는 단점을 보완 하였다. 진폭 면에서는 펄스 진폭의 다양성을 주기 위해 네 가지 경우에 대한 부호비트를 설정 하였다. 결과적으로 코드북의 크기가 증가 되었으나 진폭의 사전 설정에 의한 검색법의 단순화로 계산량의 증가 없이 성능을 향상시켰다. 제안된 알고리즘에 대한 음질 검사를 실시한 결과 주관적 음질에서 MP-MLQ와 차이가 없었으며 MNRU Q=15dB에서 제안된 알고리즘을 이용한 부호화기는 약간의 음질 저하를 느낄 수 있었다[7].

3-3. 무성음 검출에 의한 전송률 감소에 관한 연구

CELP형 보코더중 G.723.1을 사용해서 전송율을 낮추기 위해 음성신호를 묵음(Silence), 무성음(Unvoice), 유성음(Voice)으로 구분하여 부호화 하는 방법을 제안 하였다.

묵음에서 무성음으로 천이 되는 구간 중 무성음 시작 프레임, 또는 무성음에서 묵음으로 천이되는 구간에서 무성음의 끝 프레임을 검출하여 전송을 감소에 적용하였다. 그 기준은 포만트 개수와 LSP 계수의 간격의 분산을 이용하고 그에 덧붙여 스펙트럼 변화도를 먼저 측정한다.

G.723.1의 5.3kbps 음성 부호화기에서 무성음 검출 알고리즘을 C-언어로 구현하여 수행하였다. 전송율을 비교하기 위해서 크게 두 가지 과정을 수행하였다. 우선 그림 3-1의 LPC analysis 후에 얻어지는 LSP를 사용하여 검출하고자 하는 무성음 구간인지 확인한다[8].

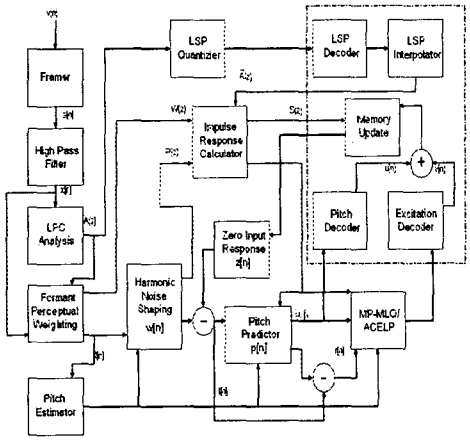

그림 3-1. 음성 부호화기의 블록도

이러한 과정에서 스펙트럼의 변화도는 20이상, 포만트 개수는 2이하, LSP간격의 분산은 20이하를 만족하는 프레임을 검출 하였다. 이러한 결정 논리에 의해 검출된 프레임이 찾고자 했던 묵음에서 무성음으로 변화하는 구간에서 무성음의 처음 부분인지 또는 묵음이나 유성음에서 무성음으로 변화하는 구간에서 무성음의 끝 부분인지 동일한 알고리즘의 Matlab 프로그램을 작성하여 확인하였다.

CELP형 부호화기의 분석과정에서 얻어지는 LSP(Line Spectrum Pair)를 이용한 성분 분리의 오차는 0.18%이하이고 전송률에 있어서 10.9%(500bps) 정도의 비트율을 낮출 수 있었으나, 음질은 0.034 정도의 열하가 발생했다.

4. 결 론

음성 전보 시스템(VTS)의 개발을 위해서는, 음성 데이터를 저렴한 하드웨어로 구현하기 위해 데이터를 많이 압축함과 동시에 저장하는 기술이 필요하다. 이러한 기술이 바로 보코더 알고리즘으로 음성 전보 시스템에서의 핵심 기술이다. CELP 계열은 낮은 전송률에서도 우수한 음질을 얻을 수 있다는 장점이 있으나 방대한 계산량 때문에 실시간 구현에 큰 어려움이 있다.

지연(delay)에 민감한 음성을 전송함에 있어서 VTS의 효과적 개발을 위해서는 CELP형 보코더들의 코드북 검색 과정에서 빚어지는 계산량을 효과적으로 줄이는 알고리즘의 개발이 이루어져야만 한다.

참 고 문 헌

[1] D.H. Yoon, Y,S,Choi,"국내외 저전송률 음성 부호화 동향" 제 13회 음성통신 및 신호처리 워크샵

[2] N. S. Jayant and P. Noll, Digital Coding of Waveform-Principles and Applications to Speech and Video, pp.220-221, Prentice-Hall, 1978.

[3] A.M. Kondoz, Digital Speech, John Willy &Sons,1994

[4] D. KIM, M. BAE, J. KIM, K. BYUN, K. HAN, H.YOO, "On a Reduction of Pitch Searching Time by Preliminary Pitch in the CELP Vocoder", J. Acoust. Soc, Korea, Vol.13, No.2E, pp.51-57, July 1994.

[5] Y.H. Park, J.C. Yang, S.M. Sohn, "On a Time Reduction of Pitch Searching by the Regular Pulse Search Technique in the CELP Vocoder," IEEE Comm. Society, Proc. of MILCOM'97, Vol.1, Nov.02-05, 1997.

[6] S.H. Hong, M.J. Bae, K.J. Byun "On a Pitch Search Technique with Correlation Characteristic of Quantization Error in CELP Type Vocoder," The Sixth Western Pacific Regional Acoustics Conference, Proceeding of WESTPRAC-VI'97, Vol.1, pp.106-111, Nov.19-21, 1997.

[7] M.J. BAE and J.D. KIM, "On a Smoothing Method by Time-Axis Control with the Zero-Crossing of Speech Signal," IEEE Communication Society, Proceeding of MILCOM'97, Vol.1, pp.-, Nov.02-05.

[8] 나덕수,정찬중,함명규,"CELP 보코더에서 무성음 검출에 의한 전송률 감소에 관한 연구",IEEE 학생논문 CONTEST 논문집, 1998.