# 음성 합성의 개선을 위한 포만트 변경에 관한 연구

이상현*, 양성일*, 권영헌**
*한양대학교 전자컴퓨터 공학부 , **한양대학교 물리학과

# Study on formant transition for improvement of speech synthesis

Sang-hyun Lee*, Sung-il Yang*, Y. Kwon**
*School of Electrical and Computer Engineering,
**Department of Physics Hanyang University
E-mail: yolong@rocketmail.com

## 요약

본 논문에서는 음성합성 과정에서 음성유닛을 연결할 때 모음의 결합부분에서 포만트의 불일치로 일어나는 부자연스러운 합성음이 발생되는 문제점을 개선하기 위해서 앞에 오는 음성 유닛과 뒤에 오는 합성 유닛의 포만트 변경에 관한 방법을 제안한다.

요즘에 연구되는 코퍼스 방식에선 에너지와 피치와 음순지속시간 등을 기준으로 유닛을 선택한 후 연결하지만, 스펙트럼의 불일치가 이루어진다. 이런 스펙트럼의 불일치는 음질의 저하를 유도한다. 그래서 앞 음성 유닛의 연결부분의 일정부분과 뒤 음성 유닛의 연결부분의 일정부분의 포만트를 천이시켜 일치시켜줌으로써 음질을 향상시켰다.

음성신호를 FFT한 후 magnitude와 phase를 분리한 후 앞 음성의 연결부분의 magnitude와 뒷 음성의 연결부분의 magnitude를 기준으로 linear interpolation한 값을 목표치로 이동하고 다시 합하여 원 신호를 복원하는 방식으로 포만트를 변경시켰다.

## 1. 서론

요즘 합성의 추세는 신호처리를 이용한 음성합성이 아닌 녹음된 음성 데이터를 유닛별로 잘라서 연결하는 연결합성 방식을 사용하고 있다. 신호처리에 의한 합성 방식은 과도한 신호왜곡으로 인해 명료도를 저하시키고 자연성 또한 크게 향상 되지 못하고 있다. 이에 반해 연결 합성 방식은 음질 면에서 좋은 성능을 가지고 있다.

연결합성방식에서도 DB의 종류에 따라 고정형과 corpus-based 방식으로 나뉜다. 고정형은 원하는 합성단위를 정한 후 단위 음성 개수를 고정하여 무한 어휘를 합성하는 방식으로 단위 음성 연결점이 고정되어있어 연결할 때 스펙트럼의 불일치가 일어나고 합성음의 명료도와 자연성에 한계가 있다. 코퍼스 방식에는 문장단위의 음성 코퍼스를 구축하고 음성 코퍼스에서 가장 적합한 단위 음성을 추출하여 신호처리 없이 이를 연결하여 합성음을 생성한다. 이 방법은 고정형의 DB보다는 자연스러운 합성음을 생성한다.

연결합성방식에서 등장하는 공통인 문제는 음성 유닛의 연결부분에서 스펙트럼의 불일치로 음질의 자연성이 떨어진다는 점이다 [1, 3, 4, 7].

이 논문에서는 연결합성방식에서 음성유닛을 접합시킬 때 발생하는 스펙트럼의 불일치를 FFT를 이용한 포만트의 천이로써 해결하고 음질의 개선을 유도하고자 한다.

## 2. 포만트의 천이

음성의 포만트는 성도를 통해 발성되는 음성신호의 공진 주파수들이다. 이 공진 주파수는 음성의 조음현상

과 조음결합과 밀접한 관계에 있을 뿐만 아니라 성대의 떨림으로 생성되는 여기신호의 피치주기가 빠르고 느림에 따라서도 영향을 받는다. 즉 피치주기가 빠르면 포만트로 비교적 높아지며 피치주기가 느리면 포만트도 비교적 낮아진다 [2].

이 논문에선 앞 음성과 뒤 음성 연결부분의 윈도우 사이즈만큼의 세그먼트에서 FFT를 구하여 magnitude와 phase정보를 나누었다. 앞 음성과 뒤 음성의 magnitude사이에 linear interpolation한 값을 삽입함으로써 포만트 천이의 자연성을 유도하였고 linear interpolation하여 구한 값을 포만트 천이의 목표치로 정하였다 [5, 6].

### 2.1 Linear Interpolation

음성에서 FFT의 magnitude를 구하면 음성의 포만트의 형태를 관찰할 수 있다. 앞에 연결할 음성과 뒤에 연결할 음성의 한 프레임에서 FFT를 구하여 magnitude값을 구하고 이들의 linear interpolation 값으로 점차적으로 변화 하는 형태를 구해야 하는데 이는 다음과 같은 형태의 식으로 구하였다.

(magnitute(FFT(앞음성)) + magnitute(FFT(뒷음성)))/2 ..(1)

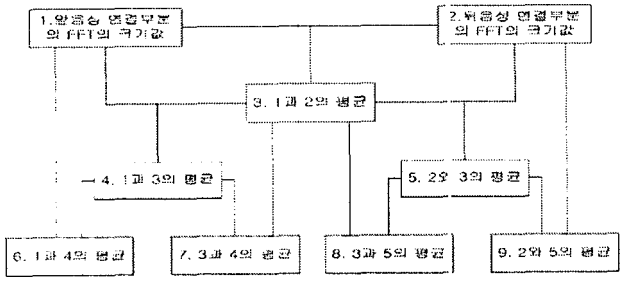


그림 1. 포만트의 목표치를 구하기 위한 과정

이런한 방식으로 평균값을 구함으로써 자연스러운 포만트의 천이가 이루어지게 할 수 있다.

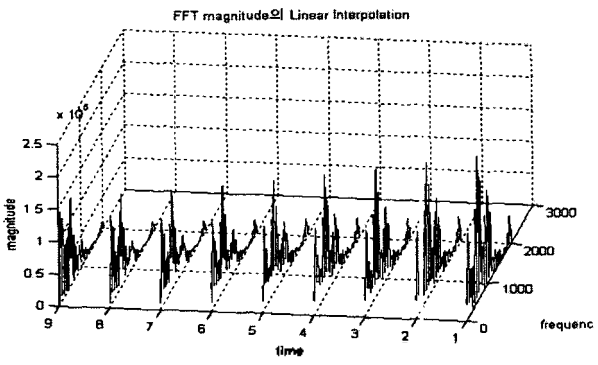


그림 2. '차'에서 추출한 '아'와 '나'에서 추출한 '아' 음성의 FFT의 magnitude의 linear interpolation한 값들의 변화

### 2.2 목표신호의 복원

FFT한 신호를 원 신호로 복원하기 위해서는 phase 정보가 필수적이다. phase정보가 달라지면 복원 시 다른 음성파형을 구하기 때문이다. 이 논문에서 원신호의 phase정보와 위에서 구한 포만트의 목표치를 이용하여 원하는 신호를 복원하였다.

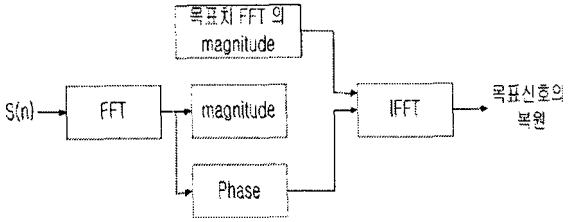


그림 3. 목표신호의 복원

## 3. 개선된 음성합성방식

음성의 포만트는 모음부분에 잘 나타나고 자음부분에서의 연결은 거의 자연성이 유지되므로 음성합성시스템에선 모음을 연결하는 방식을 택하였다.

일반적으로 포만트가 안정된 부분에서 세그먼트가 이루어져야하고 이를 이용하여 포만트 천이의 목표치를 구해야 한다.


그림 4. 합성 시스템의 간략도

발화문 생성의 과정은 한글 문자열 입력되면 여러가지 발음 규칙에 의하여 발음열로 변화되고, 이 변환된 발음열은 음성합성을 위한 입력열이 된다. (예 : 신을 신다. → 시늘 신따). 발화문이 생성되었으면 DB에서 원하는 음성유닛을 찾고 이를 연결하여 합성음을 생성한다 [8].

연결하는 과정에 앞에서 제시한 포만트 천이 방식을 이용하였다.

그림 5. '차'에서 추출한 '아'와 '나'에서 추출한 '아' 음성의 연결 후의 spectrogram.

그림 6. '차'에서 추출한 '아'와 '나'에서 추출한 '아' 음성을 위의 천이 방식으로 변경 후의 spectrogram.

Step 1 . 앞선 모음의 후반부에서 변경할 부분의 앞 프레임을 잘라서 FFT의 magnitude를 구하고 연쇄될 모음의 전반부에서 변경할 부분의 뒷 프레임을 잘라서 FFT의 magnitude를 구한다.

Step 2. 앞 음성에서 구한 magnitude와 뒤 음성에서 구한 magnitude를 가지고 linear Interpolation으로 포만트의 변경할 목표치들을 구한다.

Step 3. 앞 음성과 뒤 음성의 변경할 프레임에서 FFT의 Phase를 구한다.

Step 4. 각 목표치의 magnitude와 phase를 이용하여 IFFT로 신호를 복원하고 필요한 부분에 필요한 샘플만큼을 잘라서 음성파형을 구한다.

Step 5. 구해진 음성파형을 연결부분의 음성파형 대신 접속한다. 각각의 음성유닛의 연결에 위의 과정을 반복한다.

이와 같은 방법을 연결합성 방식에서 음성 유닛의 연결부분에 사용함으로써 자연스러운 포만트의 흐름을 유도 하였다. phase정보에 따라 음성파형의 모양이 바뀌므로 변경하고자 하는 부분의 phase정보로써 음성파형의 변경을 최소화 시켰고 FFT magnitude의 값을 변화 시켜 주면서 주파수 성분들의 변화를 자연스럽게 유도하였다.

## 4. 실험 및 결과

본장에선 위에서 제시한 음성합성 방법을 이용한 합성음과 일반적인 합성방식을 이용한 합성음을 비교 검토한다.

### 4.1 실험과정

표1. 실험조건

| | |
|---|---|
| Speech Data | 11㎑ |
| Window function | rectangular |
| Window size | magnitude부분 : 128samples |
| | phase부분 : 256samples |
| 각 magnitude에 복원할 sample 수 | 8 samples |
| FFT size | 1024*8 |

실험을 위해서 일반인 남성의 음성을 wav로 녹음하였으며 11,025㎑의 샘플링 주파수와 16비트로 양자화하여 저장하였다. 합성유닛은 diphone을 이용하였고 음성세그먼트는 여러 문장을 자연스럽게 발음한 음성에서 추출하였다.

실험문장은 '아 땅에 태어났다', '안녕하세요', '그 아이는 착하다'를 선택하여 일반적인 합성방식과 제안한 포만트의 변경을 이용한 합성방식을 이용하여 비교하였다.

### 4.2 결 과

Diphone을 이용한 음성합성에서 위에서 선정된 3개의 문장에 대하여 합성한 결과 뛰어난 음질의 개선은 이루어지지 않았으나 연결부분에서 비교적 부드러운 합성음성을 얻을 수 있었다. 하지만 유음이나 포만트가 일정하지 않은 음성의 경우 목표치가 불안정하여 약간의 음성의 변조가 생기었다. 이는 앞으로 개선해야 할 부분이라고 생각된다.

## 5. 결 론

본 연구에서는 일반적인 연결합성방식에서 음성 유닛을 연결할 때 연속으로 발성될 부분이 단독 발성된 음절의 연결로 구성되어 부자연스러운 합성음이 되는 것을 개선하기 위해서 음성신호의 FFT와 liner interpolation을 이용하여 연결부분의 포만트의 이동을 자연스럽게 해주는 방법을 제안하였다.

일반적인 다이폰의 음성합성 방식과 비교한 결과 비교적 자연스러운 합성음성을 얻을 수 있었다.

## 6. 참고문헌


[1] X. Huang, A Acero, J. Adcock, "Whistler : a Trainable text-to-speech system" . proceedings of Internatinal Conference on spoken Language Processing, Philadelpia, vol. 4 ,pp.2387~2390,1996

[2] 이기영, 최창석, "한국어 반음절단위 규칙합성의 개선을 위한 포만트천이의 변경규칙" 한국음향학회 제15권 제4호 pp. 98-104, 1996

[3] A. Ferencz, Ho-Eun song, 'Triphone-based Implementation of the Korean Hansori Trainable Text-to-speech Synthesizer" , Proceeding of ICSP 2001 August 22~24, 2001 Daejon, Korea.

[4] 김재홍, 조관선, 이철희 "연속음성으로부터 추출한 CVC 음성세그먼트 기반의 음성합성" 한국음향학회지 제 18권 제7호 (1999)

[5] Joho G. Proakis, Dimitris G. Manolakis, "Digital Signal Processing, principles, Algorithms, and Applications" , Third edition.

[6] Sanjit K. , Mitra, "Digital Signal Processing , A Computer-Based Approach" ,

[7] 권철홍 "음성인식 및 합성기술", IDEC (2000. 11.)

[8] 소병주 "음성합성을 위한 발화문 선정". 원광대학교 석사 학위 논문(1998).