

WSOLA를 이용한 동영상 미세배속 재생 서비스에 대한 콘텐츠별 배속 선호도 분석 연구

김이길*

요약

빠르게 발전하는 IT환경 속에서 이제 동영상 콘텐츠는 TV를 통한 일방 시청이 아니라 언제 어디서든 다양한 단말에서 볼 수 있는 VOD (Video on Demand) 형태로 발전하고 있다. 이러한 동영상 시청형태의 변화는 디지털이란 특성 때문에 동영상의 재생 속도 또한 다양하게 조절할 수 있다는 부가적인 장점을 사용자에게 제공 한다. 지루하고 따분한 동영상 콘텐츠는 빠르게 돌려보고 흥미 있는 장면은 느리게 천천히 볼 수 있는 동영상 미세배속 재생 기능은 오늘날 다양한 동영상 플레이어에서 제공되고 있다. 동영상 미세배속 재생 시 동영상 콘텐츠 내용의 정확한 이해를 위해서는 시각정보 못지않게 음성정보 청취가 중요한데 정상속도 보다 빠르거나 느린 재생 시 발생하는 음성의 왜곡을 줄이기 위한 음성미세배속 기술들이 음성처리 분야에서 꾸준히 발전되어 왔다. 본 논문에서는 이중 WSOLA와 같은 우수한 음성미세배속 알고리즘에 대해 알아보고 동영상 시청 시 이러한 기능 제공이 실제 얼마나 사용자 니즈(needs)에 부합하는 지 분석해보고자 한다. 특히, 동영상 콘텐츠를 사용자의 콘텐츠 소비 목적에 따라 종류별로 구분하여 재생 배속의 선호도를 조사하고 그 결과를 분석해 봄으로써 동영상 미세배속 기능 제공시 콘텐츠별 소비 목적에 맞게 재생 배속을 제공하는 것이 필요하다는 것을 제안하고자 한다.

키워드 : 멀티미디어 비디오 서비스, 동영상 배속재생 서비스, 콘텐츠 재생배속 선호도, WSOLA

A Study about the Users's Preferred Playing Speeds on Categorized Video Content using WSOLA method

I-Gil Kim*

Abstract

In a fast-paced information technology environment, consumption of video content is changing from one-way television viewing to VOD (Video on Demand) playing anywhere, anytime, on any device. This video-watching trend gives additional importance to videos with fine-speed-control, in addition to the strength of the digital video signal. Currently, many video players provide a fine-speed-control function which can speed up the video to skip a boring part, or slow it down to focus on an exciting scene. The audio information is just as important as the visual information for understanding the content of the speed-controlled video. Thus, a number of algorithms for fine-speed-control video-playing technologies have been proposed to solve the pitch distortion in the audio-processing area.

In this study, well-known techniques for prosodic modification of speech signals, WSOLA (Waveform-Similarity-Based Overlap-Add), have been applied to analyze users' needs for fine-speed-control video playing. By surveying the users' preferred speeds on categorized video content and analyzing the results, this paper proposes that various fine-speed adjustments are needed to accommodate users' preferred video consumption.

Keywords : Multimedia Video Service, Video Trick Play Service, Video Fine Speed Control, Preferred Content Playing Speed, WSOLA

※ Corresponding Author : I-Gil Kim
Received : March 13, 2015

Revised : April 20, 2015
Accepted : April 30, 2015
* KT Institute of Convergence Technology

1. 서론

일반적으로 음성의 재생속도를 빠르게 하면 남성의 목소리 톤(tone)이 여성처럼 가늘게 바뀌고 반대로 속도를 느리게 하면 여성의 목소리가 남성처럼 굵은 저음으로 변하는 현상이 일어난다. 이는 음성의 음높이(pitch)와 음색(timbre)이 왜곡되기 때문이다. 사용자의 이해를 방해하는 이러한 음성의 왜곡현상을 줄이기 위해서 목소리의 톤은 그대로 유지하면서 화자가 단지 말을 빨리하거나 느리게 하는 것처럼 느끼게 해주는 기술이 필요하게 되었고 음성처리 분야에서는 이러한 시도가 과거부터 꾸준히 연구되어왔다. 대표적으로 시간축 변환(time-scale modification : TSM)이라 불리는 이러한 음성 배속재생기술은 크게 주파수영역(frequency domain)에서 음성을 처리하는 phase vocoder[1][2]와 시간영역(time domain)에서 처리하는 중첩가산법(OLA : Overlap and Add) 계열방식[3][4][5]이 대표적인 방법이다. 이 중 OLA계열의 음색 동기 중첩가산법(PSOLA:Pitch Synchronous-based Overlap-Add)과 파형유사도 기반 중첩가산법(WSOLA : Waveform Similarity-based Overlap-Add)등이 좋은 성능으로 많이 쓰이고 있는데 최근에는 PSOLA보다 WSOLA방식이 많이 쓰이고 있다. 이것은 다른 음성배속처리 기술에 비해 그 복잡성이 낮고 왜곡처리 성능이 높아서 상대적으로 컴퓨팅 파워가 낮은 모바일 단말에 구현하기 쉬울 뿐 아니라 실시간성이 요구 되는 대부분의 동영상 서비스에 적합하기 때문이다. 이런 이유로 WSOLA를 이용한 음악배속처리[6], 실시간 동영상의 오디오 배속처리[7], 안드로이드 단말에서의 성능비교[8]등 WSOLA는 음성배속처리의 대표적인 알고리즘으로 지금까지 꾸준히 연구되고 있다.

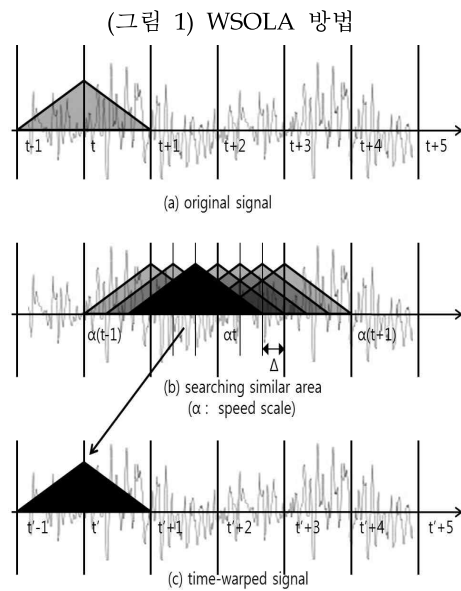
본 논문에서는 WSOLA를 이용한 동영상 미세배속 서비스를 기술적 관점과 서비스 제공 관점에서 분석해 보고자 하는데 먼저 기술적 관점에서는 현재 배속서비스에 많이 이용되고 있는 WSOLA 알고리즘에 대한 소개 와 그 성능을 평가를 서비스 관점에서는 이러한 기술이 서비스

로 제공되었을 때 사용자의 실제 선호배속을 조사, 분석하여 그 의미를 파악해 보고자 한다.

2. WSOLA기반 동영상 음성배속처리 구현

2.1 WSOLA 미세배속 구현원리

음성미세배속 시간축 변환 기술의 대표적인 예인 PSOLA, WSOLA는 주파수 영역에서의 phase vocoder 보다 상대적으로 적은 계산 량으로도 우수한 성능을 보여줌으로써 동영상 미세배속 기능의 음성처리기술로 많이 쓰이고 있다. 이 중 PSOLA는 음성을 pitch기준으로 늘이고 줄이는 방식으로 먼저 오디오 신호의 pitch를 찾아내면 그 주변을 일정범위 크기의 윈도우로 지정하고 주위 여러 개의 윈도우 안에서의 피크(peak)점끼리 겹치도록 복사 하는 방식이다 [9][10]. 이 방식은 음질은 뛰어나나 윈도우를 탐색하는 과정에서의 복잡도 증가로 인하여 실시간 처리에 어려움이 있어 주로 오프라인 환경에서의 음성합성 시에 주로 이용된다. 따라서 동영상 스트리밍 서비스처럼 실시간 음성정보 처리가 중요한 동영상 미세배속 재생 서비스의 경우는 WSOLA를 이용하는 것이 더 유리하다.

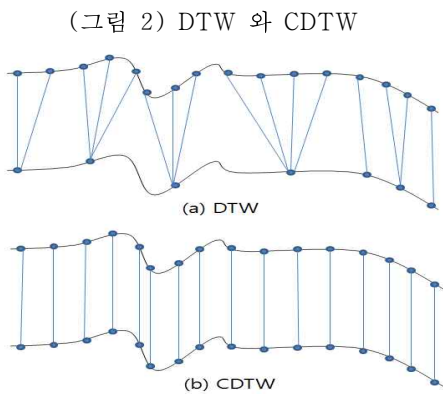


(Figure 1) WSOLA method

WSOLA방식은 시간축 상에서 중첩영역을 처리하는 방식이라는 점에서는 PSOLA와 유사하나 중첩영역을 찾아내는 방법에서 차이가 있다 [11][12]. PSOLA가 음높이인 pitch를 기준으로 유사구간을 찾는 반면 WSOLA는 일정범위의 오디오 신호의 윈도우간의 유사도(correlation)가 높은 구간을 찾는 방식으로 중첩영역을 찾아낸다. 일반적인 WSOLA 방식의 유사도 구간 검색 방법은 (그림 1)과 같다. 시간축 상의 일정 구간 즉 $t-1$ 에서 $t+1$ 까지의 오디오 신호를 (그림 1)의 (a)와 배속음성 스케일 a 만큼 증가/축소된 구간 즉 $a(t-1)$ 에서 $a(t+1)$ 사이 구간의 오디오 신호 중에서 유사도가 가장 높은 구간 (그림 1)의 (b)를 잘라내어서 재구성하면 (그림 1)의 (c)처럼 배속에 맞게 스케일링 된 오디오 신호를 만들어 낼 수 있다. 이러한 원리로 본 논문에서 동영상 선호배속 조사를 위하여 구현한 WSOLA 알고리즘은 음성신호의 반복 중첩되는 구간을 시간축 상에서 매끄럽게 늘리거나 줄여서 이어 붙이는 방식으로 음성의 왜곡 없는 배속재생이 가능하다.

2.2 WSOLA 배속 음성처리 성능평가

WSOLA 음성배속처리 알고리즘의 성능 수준은 배속 처리된 오디오 신호와 원 신호간의 파형 유사성을 비교하는 방식으로 가늠할 수 있는데 이는 비교 대상인 두 오디오 신호간의 유사성 정도를 나타내는 CDTW(Continuous Dynamic Time Warping)을 통하여 알 수 있다 [13][14].

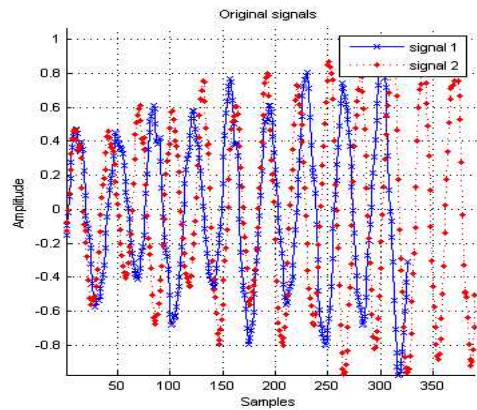


(Figure 2) DTW vs. CDTW

일반적으로 재생시간이 다른 두 개의 오디오 신호의 유사성 비교는 DTW(Dynamic Time Warping)[15][16]을 이용하는데 (그림 2)의 (a)처럼 두 개의 신호간의 모든 대응점 거리(all pair-wise distance)의 합이 적을수록 두 신호가 유사하다고 할 수 있다. 하지만 DTW의 경우 샘플링이 있는 대응점에서만 거리를 구하기 때문에 오디오와 같이 연속적인 신호의 경우에는 (그림 2)의 (b)처럼 시간축 상에서 균등한(uniform)한 간격으로 인터폴레이션(interpolation)한 샘플링 간격으로 먼저 만들어 주고 거리를 계산하는 CDTW가 더 좋은 파형 유사도 비교 성능을 보여준다.

본 논문에 사용된 WSOLA의 성능측정을 위해서 먼저 선호도 조사에 사용될 한 동영상의 오디오신호를 영상으로부터 분리한 후 WSOLA를 이용하여 고배속(이 경우 1.2배속) 변속 처리해보면 (그림 3)과 같이 고배속 처리된 signal 1이 원래 오디오 신호인 signal 2에 비해 x좌표 시간축 상에서 좌우로 압축된 것처럼 보인다. 이 상태에서 두 신호간의 유사도를 DTW 값을 구하여 알 수 있으나 (그림 4)와 같이 signal 1과 signal 2를 시간축 상의 같은 길이로 워핑(warping)하여 비교한 CDTW 값을 통하여 보다 정확한 유사도를 구할 수 있다.

(그림 3) DTW

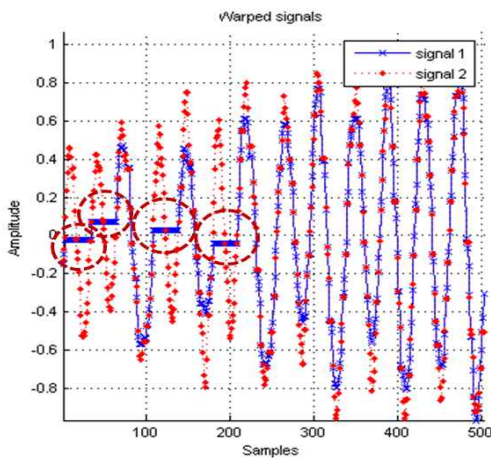


(Figure 3) DTW

또한 이러한 파형유사도 측정과정에서 WSOLA에 의한 중첩가산 과정이 제대로 일어났는지도 한번 확인해 볼 필요가 있다. (그림 4)에

서 원으로 표기된 4개 지점은 배속 처리된 signal 1 이 WSOLA의 유사도 구간 계산에 의한 중첩가산처리 되고 있음을 보여 주고 있는데 이는 오디오 신호를 시간축 상에서 자르고 이어 붙인 영역이 수평선 형태로 나타나고 있으므로 알 수 있다. 이 경우 그래프 상에서는 신호가 연속(continuous)하지 않은 것처럼 보이지만 워핑 전의 오디오 신호에서 앞 뒤 주기의 유사성이 높은 구간이 시간축 상에서 매끄럽게 이어 붙여졌기 때문에 일반적인 사람의 귀로 청취 할 때는 pitch의 변화 없이 단지 화자가 빠르게 말하는 것처럼 들리게 되는 것이다.

(그림 4) CDTW



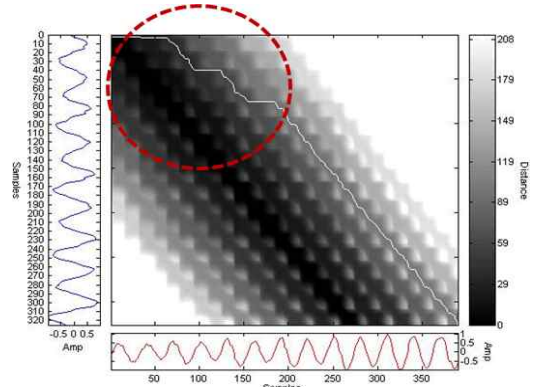
(Figure 4) CDTW

이처럼 WSOLA에 의한 중첩가산이 제대로 동작하고 있다면 그 성능 수준은 정량적 지표인 CDTW를 분석해 봄으로써 알 수 있는데 일반적으로 두 오디오 신호간의 파형 대응점의 거리의 합인 전체구간 CDTW의 축적 값이 0에 가까울수록 그래프 상에서는 그 값이 사선형태(diagonal)에 가까울수록 유사도가 높다고 할 수 있다.

앞서 1.2배속 변속처리 된 오디오의 경우 CDTW의 전체 축적 값은 17.8이고 이는 (그림 5)와 같이 그래프로 표시 할 수 있다. 이를 좀 더 자세히 분석해 보면 원으로 표기된 지점에서 사선형태의 그래프 모양이 부분적으로 이탈되고 있는데 이것은 (그림 4)에서와 같이 4개 지점에

서 발생된 WSOLA 중첩가산에 때문인 것으로 추정된다. 하지만 전체적인 그래프 값이 검은색 음영지역 안에 들어오고 있는데 이것은 두 신호의 파형의 유사도가 높다는 것을 의미한다. 따라서 본 논문의 미세배속 선호도 조사에서 사용된 WSOLA 알고리즘 구현 수준은 유사도 구간별 중첩가산이 제대로 이루어지고 있으며 파형의 유사도 또한 높은 성능을 보이고 있다는 것을 알 수 있다.

(그림 5) CDTW



(그림 5) CDTW (17.8)

3. 콘텐츠별 미세배속 선호도 조사

3.1 미세배속 선호도 조사관련 선행조사

2000년대 후반부터 국내외 인터넷 VOD동영상 스트리밍 서비스에서 선보이기 시작한 동영상 미세배속 기능은 국내의 경우 2008년 LG의 IPTV를 시작으로 2009년에 KT IPTV에서도 동영상 시청 시 음성미세배속 기능이 제공되기 시작하였다. 현재 왜곡이 없는 음성으로 동영상 배속 시청이 가능한 미세배속 재생은 IPTV, 웹, 모바일 환경에서 VOD 동영상서비스의 중요한 부가기능 중 하나로 서비스 되고 있지만 이러한 서비스에서 실제 소비자의 원하는 배속 선호도에 대한 조사는 국, 내외를 통틀어 거의 찾아 볼 수 없다. 이는 왜곡이 없는 음성배속처리 기술은 90년대부터 꾸준히 연구되어 왔지만 음성만을 이용한 배속처리 서비스에서 부분적으로 이용되어왔고 동영상 배속기능은 시청목적이 아닌 단지 구간 탐색목적으로 오디오는 대부분 무음처

리 되었기 때문에 이러한 서비스에 대한 사용자 니즈 관련 선행조사는 상대적으로 많이 부족했던 것으로 보인다. 다만 단편적인 자료긴 하지만 동영상 콘텐츠 재생 시 선호배속에 관한 설문조사가 2009년 국내의 한 인터넷 동영상 강의업체에 의해 실시된 적이 있었다[17]. 총 721명 참가한 이 설문조사에서 <표 1>과 같이 36%의 가장 많은 참가자가 1.2배속을 선호한다고 답하였는데 이는 동영상 강의라는 특성상 정배속(1X) 재생보다는 단시간 내에 많은 내용을 빠르게 학습할 수 있는 1.2배속(1.2X)을 선호한 것으로 보인다. 하지만 이 결과만 가지고는 강의라는 콘텐츠의 특성과 다른 분야의 동영상의 경우 일반적인 사용자들의 재생배속 선호도를 쉽게 예측할 수는 없을 것이다. 왜냐하면 엔터테인먼트, 어학교육, 실기강좌 등 오늘날의 동영상 콘텐츠는 각기 그 시청목적이 다르고 그에 따른 시청 특성이 다양하기 때문이다. 따라서 본 논문에서는 실제 미세배속서비스가 가능한 VOD 형태의 동영상들을 특성에 따라 분류하고 그 종류별 선호재생배속을 조사 분석해 봄으로써 오늘날 다양한 단말 환경에서의 제공되고 있는 동영상 미세배속서비스에 대한 보다 정확한 사용자의 니즈를 파악 해보고자 한다.

<표 1> 인터넷 강의 선호배속 조사결과

Preferred Speed	Count of Participants	% of Total
1X	178	24.68
1.2X	260	36.06
1.4X	156	21.63
over 1.6X	127	17.61

<Table 1> Result for the Preferred Playing Speed on an Online Lecture

3.2 배속 선호도 조사 방법 및 절차

이번 선호도 조사는 과거 인터넷업체에서 실시되었던 선호배속조사[14]와는 그 조사 방법 및 대상을 달리하여 좀 더 심도 있는 결과를 도출하고자 하였다. 이는 과거 설문조사가 콘텐츠를 제공하지 않고 단순히 선호배속에 대한 4지선다형의 단답형 이였기에 손쉽게 많은 대상자를 통한 설문을 할 수 있는 반면 실제 동영상 콘텐츠 장면의 내용과 무관한 개인의 단순 재생선호배속에 대한 결과여서 실제 콘텐츠 특성에 따른

좀 더 정확한 선호배속을 알기에는 부족하기 때문이다. 따라서 본 설문조사 시에는 동영상 콘텐츠를 같이 제공하였고 제공 동영상 콘텐츠를 특성에 따라 분류하여 다양 배속재생에서의 선호도를 조사하고 그 이유에 대한 피드백을 받아서 분석함으로써 콘텐츠 특성과 배속선호도의 관계를 알아보려고 하였다.

오늘날 온라인 동영상 서비스는 젊은 층을 중심으로 많이 소비되고 있는데[18] 이번 선호도 조사에 사용된 동영상콘텐츠 분야는 현재 IPTV와 모바일 단말에서 실제 VOD형태로 많이 소비되고 있으며 그 시청목적이 각기 다르다고 판단되는 콘텐츠들을 선정하였다. 이는 IPTV사용자 143명을 대상으로 사전 설문조사를 한 결과를 바탕으로 IPTV 콘텐츠 중 배속시청 서비스를 이용할 것 같은 콘텐츠를 6개로 선별 분류하였는데 그 동영상 콘텐츠는 엔터테인먼트의 드라마, 게임 그리고 교육목적의 어학, 통기타강좌, 댄스강좌, 골프강좌 총 6개 분야이다. 그리고 각 테스트 동영상은 실제 IPTV에서 서비스 되고 있는 콘텐츠를 이용하여 종류별로 음성정보와 화면정보가 둘 다 필요한 장면만을 선택하였고 전체 3분 정도의 샘플 영상을 캡처한 후 (그림 5)와 같이 WSOLA 기술로 구현된 음성 왜곡이 없는 미세배속 재생기능의 동영상플레이어를 선호배속을 조사하는데 사용하였다.

실제 설문 조사대상은 사전 동영상 카테고리 선별조사에 참여한 143명의 대상자 중 참가를 희망한 24명을 대상으로 하여 동영상 콘텐츠별 재생선호배속 조사를 실시하였다. 각 참가자는 6개 분야의 동영상 콘텐츠 각각에 대하여 가장 선호하는 배속과 2, 3순위의 선호배속 등 총 3개의 선호 배속을 차례로 선택 할 수 있게 하여서 총 6개의 콘텐츠에 대하여 전체 18개의 선호배속을 조사하는 방식으로 진행하였다. 이번 선호도 조사에 사용된 일부 콘텐츠가 상용이기 때문에 누출방지를 위한 서약 등의 절차적인 문제와 플레이어와 콘텐츠 제공 등 방법상의 제약으로 실제 설문 모집단은 규모는 제한되어 상대적으로 크지 않았지만 참가대상자를 실제 IPTV시청자 중 배속재생에 관심이 있는 시청자를 대상으로 하였고 실험목적에 따라 분류된 다양한 동영상 콘텐츠들에 대한 선호배속들과 그 이유를 피드백 받음으로써 과거 단순 단답형 배속 선호

도 조사보다 좀 더 깊이 있는 결과를 얻고자 하였다.

(그림 5) WSOLA 적용 미세배속 재생기



(Figure 5) WSOLA implemented Video Player

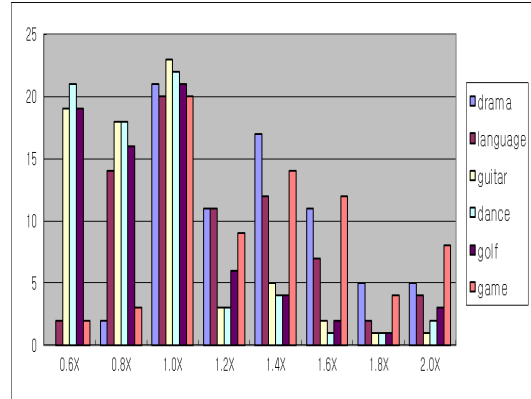
3.3 선호도 조사 결과 및 분석

전반적인 선호도 조사 결과의 흐름을 간략히 살펴보면 전체 6종류의 동영상 콘텐츠에 대한 사용자 선호도는 (그림 6)과 같은 결과를 보이는데 이 그림에서 각 그래프는 6개 콘텐츠에 대한 1, 2, 3순위 선호배속의 총합을 나타낸다. 이 결과만 보면 동영상 종류 특성과 상관없이 사용자의 선호배속은 저배속과 고배속 양쪽으로 다양하게 분포하고 있음을 알 수 있다. 하지만 이를 좀 더 자세히 들여다보면 동영상의 시청목적과 시청특성 별로 그 선호배속이 다를 수 있고 이제 그 결과를 콘텐츠별로 하나하나 분석하여 보고자 한다.

먼저 빠른 미세배속을 선호 하는 콘텐츠의 경우인데 (그림 7)과 같이 드라마와 게임 콘텐츠는 일반적으로 정배속(1순위)을 가장 선호했지만 이에 못지않게 1.4배속(2순위) 또는 1.6배속(3순위) 까지도 선호하는 특징을 보였다. 이는 왜곡이 없는 음성정보를 시각 정보와 더불어 제공한다면 VOD 사용자는 1.6배속 시청도 많이 이용할 수 있다는 것을 보여준다. 현재 대부분의 IPTV 배속시청 기능이 1.2배속까지만 미세배속 음성을 지원하는 것을 보면 좀 더 높은 미세배속에 대한 니즈도 존재 하고 있음을 여기서 확인 할 수 있다. 참고로 참여자들의 부가 의견에 따르면 드라마는 농친 회에 대한 전체적인 즐거리를 따라

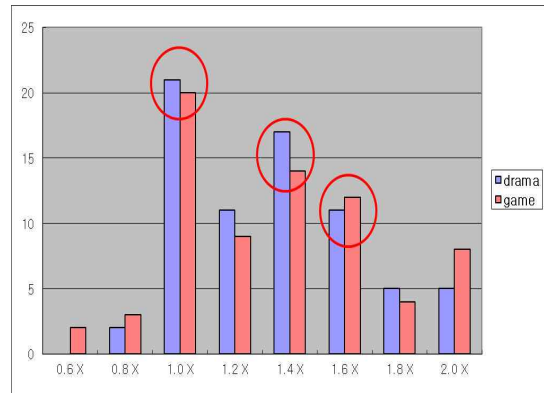
가기 위해 장면 장면의 의미 정보보다는 전체적인 줄거리 파악을 위해 빠르게 보기를 원했고 게임의 경우도 전체적인 게임의 진행과 승부의 흐름을 파악하기 위해 빠른 미세배속을 선호한다고 하였다. 이는 다시보기라는 특성상 이런 종류의 콘텐츠는 상대적으로 정배속보다 빠르게 보고자 하는 니즈가 많았던 것으로 해석 될 수 있다.

(그림 6) 선호배속 결과



(Figure 6) Results of Preferred Speed

(그림 7) 빠른 배속 선호

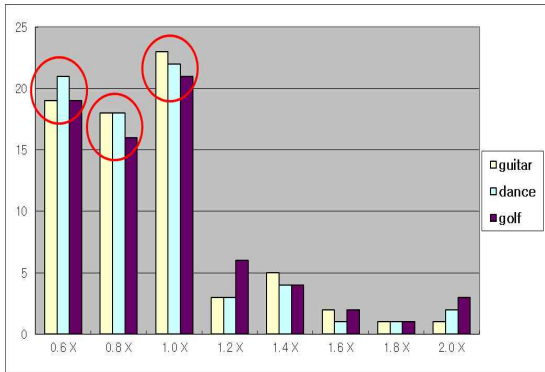


(Figure 7) Fast Speed Preferred

반면 느린 미세배속을 선호 하는 콘텐츠도 있었는데 (그림 8)과 같이 통기타, 댄스, 골프강좌와 같은 실기성 강좌 콘텐츠들은 상대적으로 느린 미세배속을 더 선호하는 현상이 두드러졌다. 정배속(1순위)외에 0.6배속(2순위), 0.8배속(3순위)과 같이 저배속을 선호한 참여자가 고배속 선

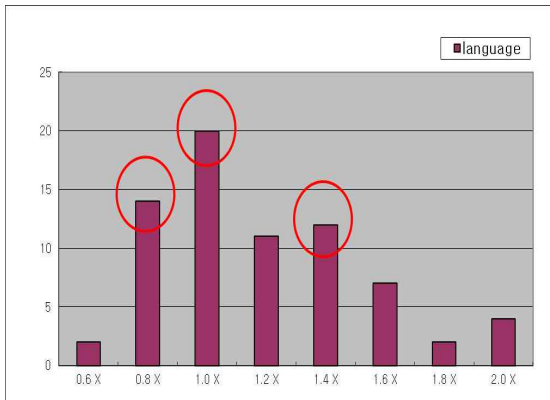
택 참여자 보다 월등히 높았다. 이는 기타연주의 손가락 움직임, 댄스의 몸동작, 골프의 스윙동작 한 장면 한 장면을 천천히 보고 따라서 학습하고자 하는 시청목적이 느린 미세배속 시청특성으로 나타난 것이라 판단된다.

(그림 8) 느린 배속 선호



(Figure 8) Slow Speed Preferred

(그림 9) 고/저배속 선호



(Figure 9) Fast and Slow Speed Preferred

그렇지만 같은 학습이라는 카테고리지만 어학 강좌 콘텐츠는 조금 다른 특성을 보였다. 실기성 위주의 콘텐츠와 달리 어학은 오디오 정보가 상대적으로 중요한 콘텐츠이다. (그림 9)에서 보는 것과 같이 정배속(1순위)외에도 0.8배속(2순위)와 1.4배속(3순위)등 저배속과 고배속 양쪽에 대한 니즈 공존하는 결과가 나왔다. 이는 동영상 콘텐츠 전반에 대한 선호 배속이라기보다는 어학이라는 특성상 특정 구간의 어려운 발음은 느린 배속으로 재생하면서 따라 하기도 하고 또 다른

구간은 빠르게 속칭하면서 자신의 리스닝 능력을 높이고자 하는 다양한 니즈가 있기 때문에 결과적으로 좀 더 다양한 미세배속 제공이 가장 필요한 콘텐츠라고 생각된다.

4. 결론

현재 음성의 왜곡 없이 저배속, 고배속 재생이 가능한 동영상스트리밍 서비스는 IPTV, 온라인 어학 사이트 등과 같이 많은 곳에서 VOD 동영상 미세배속 서비스라는 이름으로 서비스 되고 있다. 이는 시간에 쫓기는 바쁜 현대인들에게 빠른 시청기능을 제공해주고 학생들에게는 천천히 따라 학습할 수 있게 느린 재생기능으로 도움을 주기도 하는 등 오늘날 사용자에게 꼭 필요한 서비스 중에 하나이다. 하지만 대부분의 음성이 지원되는 동영상 미세배속 서비스는 콘텐츠의 특성과 상관없이 사업자에 의해 일방적으로 정해진 몇 가지 배속으로만 서비스가 제공되고 있는 것이 현실이다.

본 논문에서는 WSOLA알고리즘을 이용하여 음성의 왜곡이 없는 음성미세배속 방법을 다양한 동영상에 적용하여 콘텐츠 종류별 사용자 선호배속을 조사하여 보았는데 그 결과 엔터테인먼트는 빠른 미세배속을 실기강좌에서는 느린 미세배속을 선호하였고 어학 같이 오디오 정보가 중요한 콘텐츠에서는 저/고배속의 좀 더 다양한 미세배속을 사용자가 선호한다는 것을 알 수 있었다. 이러한 소비자의 니즈를 수용하여 사업자들은 동영상 미세배속 서비스의 배속 선택을 콘텐츠 특성에 맞게 좀 더 다양하게 제공한다면 서비스 만족도는 더욱 높아 질 것이다.

References

[1] J. Laroche and M. Dolson, "Improved phase vocoder time-scale modification of audio," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 3, pp. 323 - 332, May 1999.

[2] D. W. Griffin and J. S. Lim, "Signal estimation from modified short time Fourier transform," *IEEE Trans. Audio, Speech, Signal Process.*, vol. ASSP-32, no. 2, pp. 236 - 243, Apr. 1984.

- [3] E. Moulines and J. Laroche, "Non-parametric techniques for pitchscale and time-scale modification of speech," *Speech Commun.*, vol. 16, no. 2, pp. 175 - 206, 1995.
- [4] E. Moulines and F. Charpentier, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Speech Commun.*, vol. 9, no. 5-6, pp. 453 - 467, 1990.
- [5] W. Verhelst, "Overlap-add methods for time-scaling of speech," *Speech Commun.*, vol. 30, no. 4, pp. 207 - 221, 2000.
- [6] Shahaf Grofit, Yizhar Lavner, "TimeScale Modification of Audio Signals Using Enhanced WSOLA With Management of Transients", *IEEE Transactions on Audio, Speech & Language Processing - TASLP*, vol. 16, no. 1, pp. 106-115, 2008
- [7] Ivan Damjanovic, Dan Barry, David Dorran, Joshua D. Reiss, "A Real-Time Framework for Video Time and Pitch Scale Modification," *IEEE Transactions on Multimedia - TMM*, vol. 12, no. 4, pp. 247-256, 2010
- [8] Wlodarczyk, M., Sekalski, P., "Evaluation of time-scale modification methods for audio signals on mobile devices with android OS", *Proceedings of the 21st International Conference on Mixed Design of Integrated Circuits & Systems (MIXDES)*, 2014
- [9] H.Valbret,E.Moulines,andJ.P.Tubach,"Voice transformation using PSOLA techniques," *Speech Communication.*, vol. 11, pp. 175 - 187, 1992.
- [10] S. Roucos and A. Wilgus, "High quality time-scale modification of speech," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Tampa, FL, Mar., pp. 493 - 496, 1985.
- [11] S. Grofit and Y. Lavner, Time-scale modification of audio signals using enhanced wsola with management of transients, *IEEE Transactions on Audio, Speech & Language Processing*, 16, pp. 106 - 115, 2008
- [12] W. Verhelst and M. Roelands, "An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing. (ICASSP)*, Minneapolis, MN, pp. 554 - 557, 1993.
- [13] A. Efrat, Q. Fan, and S. Venkatasubramanian. Curve matching, time warping, and light fields: New algorithms for computing similarity between curves. *J. Mathematic Imaging and Vision*, 2007.
- [14] M. Munich and P. Perona, "Continuous dynamic time warping for translation-invariant curve alignment with applications to signature verification," in *International Conference on Computer Vision (ICCV)*, pp. 108 - 115, 1999.
- [15] K. Huang and H. Yan. On-line signature verification based on dynamic segmentation and global and local matching. *Optical Engineering*, 34(12):3480 - 3487, 1995.
- [16] R. Martens and L. Claesen. On-line signature verification by dynamic time-warping. In *Proc. 13th Int. Conf. Pattern Recognition*, pages 38 - 42, 1996.
- [17] http://www.g-school.co.kr/community/pollEnd.js?poll_code=2009030400001
- [18] Sun-jin Kim, The present and prospect of Online Video, Music service and Media Usage, *Journal of Digital Contents Society.*, vol. 16, pp.137-144, 2015



김 이 길

2002년 : University of Southern California, Electrical Engineering (M.S.)

2008년 : University of Florida, Electrical and Computer Engineering (E.E.)

2003년~2005년: 삼성전자 Digital Media 선임연구원
 2008년~현재: KT 융합기술원 책임연구원
 관심분야 : 영상처리(Image Processing), 가상현실(Virtual Reality), 증강현실(Augmented Reality), 실감미디어(Realistic Media), 게임 프로그래밍(Game Programming) 등