

워터마킹을 이용한 동적캐릭터

A dynamic character using watermarking technique

박 귀 태, 김 갑 일, 손 영 익

(Kyi Tae Park, Kab Il Kim, and Young Ik Son)

명지대학교 차세대 전력연구센터(E-mail : superz@hanmail.net, {kkl, sonyi}@mju.ac.kr)

Abstract : 본 논문에서는 워터마킹을 응용한 동적 캐릭터 기술을 제안한다. 오디오 신호에 임의의 동작을 위한 일련의 동작코드를 들리지 않게 삽입하면, 이 캐릭터는 선행 처리된 오디오 파일이 재생될 때 마이크를 통해 얻은 소리를 처리하고 은닉된 코드를 추출함으로써 그 코드에 지정된 행동을 할 수 있다. 예를 들어 로봇이 동작코드가 은닉된 음악에 맞추어 지시된 춤을 출 수 있는 것이다. 이를 위해 우리는 워터마킹 기술을 적절히 응용하였으며, 제안된 기술은 공기(air)를 매질로 삼은 아날로그 채널에서 발생하는 잡음, 로봇과 음원의 거리에 따른 음의 감쇄, 그리고 동기화 등과 같은 문제들을 해결하였다. 여러 가지 상황에 따른 제안된 기술의 실험을 통해 제안된 기술의 성능을 입증하였다.

Keywords : character, robot, watermark, watermarking, audio, wireless

1. 서론

최근 로봇산업은 인공지능이나 음성인식과 같은 첨단 기술의 발전과 더불어 지능적 행동을 발현할 수 있는 동적 캐릭터들의 개발을 촉진하였다. 소니(SONY)의 AIBO를 선두로 오무론(Omron)의 네코로(NeCoRo), 반다이(Bandai)의 BN-1등과 같은 애완로봇은 실제 애완동물과 같이 느끼고 행동할 수 있도록 지능 프로그램이 내장되었으며 시장에 진출해 고부가가치 산업을 형성하였다. 특히 이 중 소니는 AIBO를 위한 TV프로그램을 제공하는데, 사용자가 이 프로그램을 AIBO와 함께 시청하면 각 장면마다 AIBO가 특정 전자음을 발생시킴으로써 살아서 감정을 느끼는 듯한 인상을 준다. 이와 유사하게 새로운 캐릭터 모델 또는 기존의 애완로봇을 위한 부가기능으로 로봇이 음악을 듣고 특정코드를 인식해 춤을 추거나 특정 행동을 할 수 있는 기술을 고안하였다.

제안된 기술은 워터마킹을 응용하였는데, 워터마킹이란 원래 디지털 콘텐츠의 저작권 보호를 위한 기술로써 저작권 정보를 영상이나 오디오 같은 디지털 콘텐츠에 사람이 인식할 수 없도록 삽입하여 소유권을 주장할 수 있게끔 해준다. 이 때 삽입되는 저작권 정보를 워터마크라 부른다. 이 기술은 어떤 정보를 은밀히 숨기고 추출하기에 용이하다는 점에서 새로운 캐릭터를 위한 기술의 토대로 삼았다.

현재까지는 제안된 캐릭터 모델에 대한 연구사태가 없지만 저작권 보호기술로써의 워터마킹에 관해서는 많은 연구들이 진척되어왔다[1-3]. 오디오 워터마킹의 선행연구를 살펴보면 크게 Spread Spectrum을 이용한 방법, 복

제신호변조를 이용한 방법, 그리고 반향삽입에 의한 방법으로 분류할 수 있다. Spread Spectrum을 이용한 연구 [4-8]는 워터마크를 나타내는 PN-sequence를 심리음향 모델로 마스킹하여 삽입함으로써 삽입된 워터마크의 비가청성과 견고성을 만족시킨다. 복제 신호(Replica signal) 변조를 이용한 기법[9]은 오디오 신호의 일부를 주파수의 위치 또는 위상 지연을 통해 약간 변형하여 워터마크로 이용한다. 반향삽입법(echo hiding)[10]은 사람의 귀가 짧은 시간간격을 두고 발생하는 반향에 둔감하다는 점을 착안, 이를 이용하여 워터마크를 삽입한다. 여기서 반향삽입법은 하나의 큰 반향만을 삽입하므로 음질의 저하를 가져오게 되는데, 다중반향삽입 기법[11-13]은 하나의 큰 반향 대신 여러 개의 작은 반향을 삽입함으로써 음질저하를 극복하였다.

제안된 캐릭터 모델은 스피커로부터 나온 음악의 소리를 공기(air)라는 무선 아날로그 채널에 실어 캐릭터에게 전송하므로 이에 따른 음의 감쇄와 많은 잡음을 동반하게 된다. 기존의 워터마킹기법들이 D/A-A/D변환 공격에 대해 우수한 성능을 보여 왔지만 이는 주로 전선을 이용한 유사 실험들으로써 단지 소량의 잡음과 데이터 누수만이 발생 할뿐, 공기를 매질로 삼은 D/A-A/D변환에 의해 발생하는 문제점들을 해결할 수 있는 워터마킹 기술이라 볼 수 없다. 그래서 우리는 이를 해결할 수 있는 방법을 모색하였고 반향삽입법이 음의 감쇄나 동기화 문제에 비교적 덜 민감하다는 것을 발견하였다. 하지만 무선 아날로그 채널에서 생기는 급격한 음의 감쇄와 많은 잡음 때문에 워터마크 검출에서 낮은 성능을 보이므로 본 논문에서는 이를 극

복하고 제안된 캐릭터에 적용할 수 있는 반향삽입법 기반의 무선 오디오 워터마킹 기술(wireless audio watermarking technique)을 제안하려 한다.

본 논문의 나머지 구성은 다음과 같다. II장에서는 제안된 기술에 대해 설명하고 III장에서는 제안된 기술의 성능을 입증할 수 있는 실험결과를 제시한다. 마지막으로 IV장에서는 향후 연구방향과 함께 결론을 맺는다.

II. 제안된 기술

반향삽입법이 가지는 중요한 특징은 워터마크 추출 시, Spread Spectrum 기법처럼 문턱치(threshold)를 두어 워터마크를 결정하는 절대비교방식이 아닌 우세한 쪽을 취하는 상대비교방식을 이용한다는 것이다. 무선 아날로그 채널에서는 음의 감쇄 때문에 워터마크 추출을 위한 문턱치 설정이 어려우며 실시간으로 워터마크를 검출해야 하므로 이것의 추정 또한 쉽지 않다. 이러한 점을 고려할 때 워터마크 검출을 위해 상대비교방식을 이용하는 것이 유리하다는 것을 알 수 있다. 이러한 상대비교방식과 함께 반향삽입법은 동기화 문제에 비교적 덜 민감하고 실시간 워터마크 검출이 가능하므로 제안된 기술을 위한 기반 알고리즘으로 적합하다.

하지만 반향삽입법이 무선 워터마킹을 위한 몇 가지 장점을 가졌음에도 불구하고 이를 바로 이용하기에는 몇 가지 취약점이 드러난다. 한 가지 문제점으로 무선 워터마킹 환경에서 스피커와 로보간의 거리에 따른 음량의 급격한 감소와 높은 잡음에 때문에 음의 SNR(Signal-to-Noise Rate)이 매우 낮아지고 워터마크 추출확률이 현저히 떨어져 수정을 위한 에러정정코드 조차 사용이 용이하지 못하다는 것이다. 우리는 이를 극복하기 위한 방법의 일환으로 코드(워터마킹에서는 워터마크)를 반복적으로 삽입하고 추출 시 다수결의 원칙을 이용해 숨겨진 코드를 결정하는 방식을 택하였다. 하지만 코드의 반복횟수 만큼 삽입할 수 있는 코드의 수는 감소하므로 2가지 밖에 표현할 수 없는 바이너리 대신 정수 코드를 사용하였다. 기존의 반향삽입법이 0과 1을 표현하기 위해 각각 δ_0 , δ_1 의 시간지연을 가지는 반향을 사용하였던 것과는 달리, 우리는 0~K의 범위를 가지는 코드를 표현하기 위해 $\delta_0 \sim \delta_K$ 의 시간지연을 가지는 반향을 설정하였다. 이것은 여러 종류의 반향을 이용하지만 하나의 정보를 표현하기 위해 한 개의 반향만을 삽입하므로 한 번에 여러 개의 작은 반향을 동시에 삽입하는 멀티반향삽입법[11-13]과는 다르다. 여러 종류의 반향을 사용하는 방법은 비록 오디오의 음질을 열화시키지만, 우리는 대다수의 사람들이 오디오의 음질보다는 동작하는 로봇에 관심이 모아지기 때문에 무시할 수 있다. 다른 한 가지 문제점은 동기화에 있다. 동기화란 코드가 삽입된 시작점을 찾는 것인데, D/A-A/D변환이 있을 시 코드가 삽입된 시작점을 알 수 없으며 그림1처럼 코드 추출을 위해 처리되는 블록과 실제 삽입된 블록이 어긋나게 된다. 따라

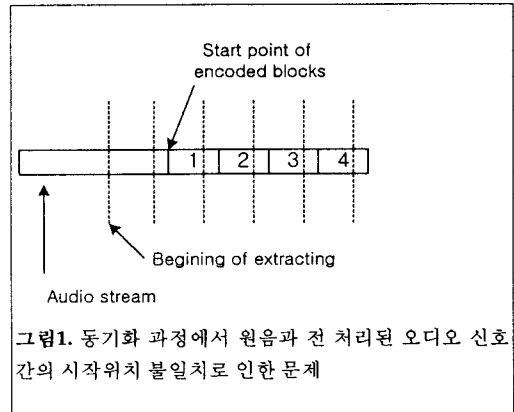


그림1. 동기화 과정에서 원음과 전 처리된 오디오 신호간의 시작위치 불일치로 인한 문제

서 하나의 처리블록에는 두 개의 서로 다른 반향이 포함되며 이는 결국 데이터의 추출 확률을 떨어뜨리는 요인으로 작용하게 된다. 그러나 위에서 언급된 것처럼 같은 코드를 반복하여 삽입할 경우 같은 반향이 연속하여 존재하므로 처리블록이 실제블록과 어긋남에 의한 피해를 최소화할 수 있다. 이를 이용해 몇 개의 연속된 처리블록에서 일정수의 특징코드(동기화 코드)가 추출되었을 때 그 부분을 코드 삽입의 시작점으로 결정하는 근사동기화 할 수 있게 된다.

이와 같은 코드 반복삽입법과 함께 이어지는 절에서 동작코드의 삽입과 추출 과정에 대해 설명한다.

1. 동작코드 삽입

우리의 기법은 동기화와 코드삽입을 위해 다음과 같은 과정을 가진다.

Step1: 우선 코드의 삽입을 위해 기존의 반향삽입법이 0과 1로만 워터마크를 표현했던 것과 달리 필요한 만큼 코드의 범위를 정의한다.

0: 부 동작 코드, 1: 동기화 코드, 2~K: 동작코드

여기서 0은 부 동작 코드로 로봇이 이 코드를 인식했을 때 아무 동작도 하지 않을 것을 지시한다. 1은 동기화를 위한 코드이며 2~K는 캐릭터의 행동을 지시하기 위한 코드이다. 코드가 0-K까지의 범위를 가지기 때문에 $\delta_0 \sim \delta_K$ 의 시간지연을 가지는 반향을 이용한다.

Step2: 하나의 동작코드를 삽입하기 위해 그림2와 같이 프레임에 몇 개의 블록을 묶어 구성한다. 그림2에서 하나의 프레임에는 동기화 코드를 위한 8개의 블록과 동작코드를 위한 5개의 블록이 있다. 블록들은 추출확률을 높이기 위해 반복삽입 되었으며 8개와 5개의 반복회수는 실험을 위해 무작위로 선정하였다.

Step3: 원하는 코드들을 모두 넣기 위해 Step2의 프레임에 필요한 만큼 삽입한다. 예를 들어 5개의 동작 시퀀스를 삽

입하려면 5개의 프레임을 차례대로 오디오 신호에 삽입하면 된다.

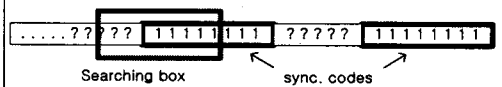
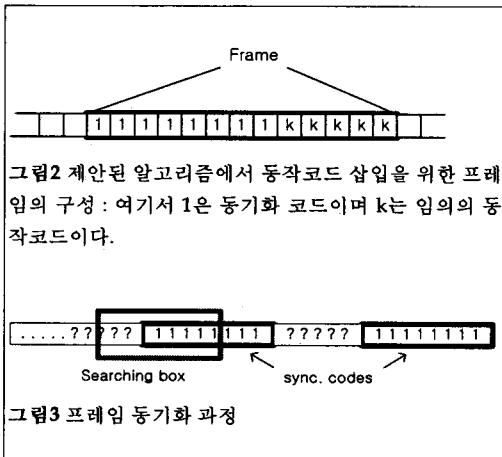
2. 동작코드 추출

동작코드를 추출하기 위해서는 프레임 동기화 후 프레임 내 동작코드를 추출하면 된다. 아래에서는 블록으로부터 코드를 추출하는 방법, 동기화, 프레임 내 동작코드 결정법에 대해 설명한다.

Step1: 각 블록 내 숨겨진 코드를 추출하는 방법은 반향삽입법과 유사하다. 여기에서도 역시 `auto-cepstrum[10]`을 이용하는데, 블록의 `auto-cepstrum` 이 $c[\delta_i]$ 라 할 때 반향삽입법에서 $c[\delta_0]$, $c[\delta_1]$ 중에서 큰 것을 취하여 그 블록의 워터마크 비트로 결정한 것처럼 여기서도 $c[\delta_0] \sim c[\delta_k]$ 중 가장 큰 것을 취하면 된다. 예를 들어 이 중 $c[\delta_2]$ 가 가장 큰 값을 가졌다면 '2'가 그 블록에서 추출된 코드인 것이다.

Step2: 프레임 동기화를 위해서는 그림3처럼 박스 안의 8개 블록에서 코드들을 추출해내야 한다. 그림3의 경우 박스 안의 8개의 블록들로부터 코드를 추출하면 3개의 '?' 코드와 5개의 동기코드를 얻는다. 만약 박스가 프레임의 시작블록부터 놓이게 된다면 8개의 동기화 코드가 모두 추출되므로 이 지점이 동기라고 판정할 수 있다. 일반적인 상황에서 '?' 표시된 블록에는 동기화 코드를 사용하지 않지만 최악의 경우 추출 에러에 의해 3개의 '?'가 모두 1로 판독되면 잘못된 동기화가 이루어지게 된다.

Step3: 동기화가 이루어졌다면 이제 동기화 코드 다음에 위치한 5개의 연속된 블록들로부터 동작코드를 추출해 다수결을 취하면 된다. 예를 들어 한 프레임에서 '4','4','2','4','4'라는 5개의 동작코드들을 추출하였다면 이 프레임의 동작코드는 다수결의 원칙에 따라 4로 결정되는 것이다. 이제 이러한 방법을 반복하여 모든 프레임의 동작코드를 찾아낼 수 있다



III. 실험결과

본 논문에서는 실험을 위해 팝송과 댄스곡 및 스피치 파일의 소리 데이터를 사용하였다. 시행된 실험에서 원음과 반향의 상대볼륨을 0.7로 고정하고 블록크기 (nfft), 프레임이 가지는 동작코드의 개수 (ncodes), 코드의 범위 (ntypes), 스피커와 로봇간의 거리 (distance)를 각기 달리 해 동작코드의 추출확률을 구해보았다.

블록크기를 달리하며 시행한 동작코드의 추출에서는 2048이상의 크기를 가질 때 비교적 신뢰할만한 결과를 얻을 수 있었다(표1). 블록크기가 512 일 때 스피치 파일의 동작코드 추출확률은 'x'로 표시되었는데 이는 확률이 너무 낮아 추출이 불가능함을 뜻한다.

스피커와 로봇간의 거리에 따른 동작코드의 추출에서는 둘 간의 거리에 비례하여 동작코드의 추출확률도 낮아짐을 보였다(표2). 이는 로봇이 음원으로부터 멀어질수록 받아들이는 음량도 작아져 음의 SNR이 감소하기 때문이라 추정할 수 있다.

프레임 내 동작코드의 개수를 달리해가며 시행한 실험에서는 프레임 내 동작코드의 개수에 비례하여 동작코드의 추출확률이 높아짐을 보인다(표3). 실험에서는 이것이 3개 이상일 때 비교적 신뢰할만한 결과를 보였다.

마지막으로 사용되는 동작코드의 범위는 동작코드 추출확률에 큰 영향을 미치지 않았다(표4). 여기서 동작코드의 범위는 총 코드 수 K에서 부 동작코드와 동기화 코드를 뺀 $K-2$ 이다.

시행된 실험에서 팝송이나 댄스곡 같이 비교적 복잡하고 정적(silence)이 없는 신호에서는 양호한 결과를 보였다. 반면 스피치 파일의 경우에는 저조한 결과를 보였으며, 제시하지는 않았지만 클래식과 같이 조용한 음악에서는 거의 추출이 불가능하였다.

IV. 결론

본고는 새로운 캐릭터 모델을 위한 기술을 제안하였다. 새로운 캐릭터 모델은 음악에 삽입된 은닉코드를 탐지하고 이에 맞추어 행동할 수 있는 능력을 가지고 있으며, 이는 실시간으로 음악을 들으며 동작코드를 추출해야 하는 무선 워터마킹 기술을 필요로 한다. 우리는 이를 위해 무선 아날로그 채널에서 은닉된 코드들을 추출할 수 있는 기술을 마련하였으며 실험을 통해 신뢰할만한 수준의 코드 추출을 할 수 있다는 것을 보였다.

비록 제안된 기술이 새로운 캐릭터 모델을 위한 소기의 목적을 달성하였지만, 각기 다른 시간지연을 가지는 반향의 사용으로 음질떨화를 초래하였다. 보다 우수한 성능을 위해 음질향상을 위한 후속연구가 이루어져야 하겠다.

표1. 블록크기에 따른 동작코드의 추출확률 : distance=1m, ncodes=5, ntypes=6, ratio=0.7

nfft	512	1024	2048	4096
pop	59.2%	91.4%	97.6%	99.7%
dance	76.7%	90.6%	99.6%	99.9%
speech	x	46.7%	68.3%	83.9%

표2. 거리에 따른 동작코드의 추출확률 : nfft=2048, ncodes=5, ntypes=6, ratio=0.7

distance	50cm	1m	2m	3m
pop	99.8%	97.6%	92.8%	84.8%
dance	99.6%	99.8%	99.6%	97.6%
speech	77.3%	68.3%	31.5%	x

표3. 프레임 내 동작코드 수에 따른 동작코드의 추출확률 : nfft=2048, distance=1m, ntypes=4, ratio=0.7

ncodes	1	3	5	7
pop	x	96.6%	97.6%	99.6%
dance	x	99.5%	99.9%	99.7%
speech	x	51.1%	68.3%	86.1%

표4. 데이터 타입 수에 따른 동작코드의 추출확률 : nfft=2048, distance=1m, ntypes=6, ratio=0.7

ntypes	2	4	6	8
pop	99.6%	97.6%	99.7%	99.7%
dance	99.6%	99.7%	99.9%	99.8%
speech	66.6%	68.3%	65.1%	63.1%

감사의글

The authors would like to thank the Korea Ministry of Science and Technology and the Korea Science and Engineering Foundation for their support through the ERC program.

참고문헌

[1] C. I. Podilchuk and E. J. Delp, "Digital watermarking: algorithms and applications," *IEEE Signal Processing Magazine*, vol. 18 no. 4, pp. 33-46, 2001.

[2] W. Bender, D. Gruhl, N. Morimoto, A. Lu, "Techniques for data hiding," *IBM System Journal*, vol. 35, no. 3 & 4, 1996.

[3] I. J. Cox et al., "Secure spread spectrum watermarking for multimedia", *IEEE Trans. on Image Processing*, vol. 6 no. 12, pp. 1673-1687, 1997.

[4] L. Boney, A. H. Tewfik, and K. N. Hamdy, "Digital watermarks for audio signals," *Proc. of IEEE Int. Conf. on Multimedia*, 17-23 Jun. 1996.

[5] D. Kirovski and H. Malvar, "Robust spread-spectrum audio watermarking," *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1345-1348, 2001.

[6] L. Zheng and A. Inoue, "Audio watermarking techniques using sinusoidal patterns based on pseudorandom sequences," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 8, pp. 801-812, 2003.

[7] S. K. Lee and Y. S. Ho, "Digital audio watermarking in the cepstrum domain," *IEEE Trans. on Consumer Electronics*, vol. 46, no. 3, pp. 744-750, 2000.

[8] A. N. Lemma, J. Apea, W. Oomen, and L. van de Kerkhof, "A temporal domain audio watermarking technique," *IEEE Trans. on Signal Processing*, vol. 51, no. 4, pp. 1088-1097, 2003.

[9] R. Petrovic, "Audio signal watermarking based on replica modulation," *Proc. of IEEE Int. Conf. on Telecommunications in Modern Satellite, Cable, and Broadcasting Service*, vol. 1, pp. 227-234, 2001.

[10] D. Gruhl, A. Lu, and W. Bender, "Echo hiding," *Proc. of IEEE Workshop on Information Hiding*, 1996, pp. 295-315.

[11] C. Xu, and et al., "Applications of digital watermarking technology in audio signals," *J. Audio Eng. Society*, vol. 47, no. 10, 1999.

[12] H. O. Oh, J. W. Seok, J. W. Hong, and D. H. Youn, "New echo embedding technique for robust and imperceptible audio watermarking," *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 1341-1344, 2001.

[13] S. W. Foo, T. H. Yeo, and D. Y. Huang, "An adaptive audio watermarking system," *Proc. of IEEE Int. Conf. on Electrical and Electronic Technology*, vol. 2, pp. 509-513, 2001.