한국어 TTS 시스템을 위한 운율구 경계 예측

전진욱 , 김한우
한양대학교 컴퓨터공학과

김동건 , 이양희
동덕여자대학교 전산정보학부

Prosodic-Boundary Prediction for Korean Text-to-Speech System

Jin-wook Chun Han Woo Kim
Dept. of Computer Engneering

Dong gun Kim Yanghee Lee
Dept. of Computer & Info. Science

Hanyang University, Dongduk Women's University
{jwchun, kimhw} @ cse.hanyang.ac.kr, {dongg, yhlee} @ dongduk.ac.kr

요 약

운율은 음성의 초분절적인 면에 연관하는 음성의 한 특성으로서 통상적으로 화자는 음성을 전달하는 과정에서 청자의 이해를 돕기 위해 운율을 사용하게 된다. 본 논문은 이러한 운율을 이루는 성분 중의 하나인 운율구의 위치 예측에 대한 성능을 향상시키는 것에 그 목적을 둔다. 한국어 운율 정보에 대한 표기 방법 중의 하나인 K-ToBI를 기반으로 하여, 운율구의 경계와 그에 대한 레벨을 Break Indices 정보로서 나타내었고, 통계학 분야에서 제안된 Support Vector Machine(SVM)을 이용하여 시스템의 예측률 향상을 꾀하였다. 기존의 방법에서 사용된 트리 기반 모델을 이용하여 한국어 운율에 가장 많은 영향을 끼치는 언어 정보들을 추출하였고 이를 실험에 적용하였다. 기존의 트리 모델과 SVM 모델에 대한 예측률을 비교한 결과, 경계 유무 정보 예측과 4단계의 레벨을 가지는 경계 정보의 예측에서 모두 본 방법이 보다 높은 예측률을 보여 주어 본 연구에서 제시한 접근법이 운율구의 경계 정보를 예측하는 데에 있어 더욱 효과적인 접근법임을 실험적으로 입증하였다.

1. 서 론

화자는 문장을 발화함에 있어 일반적으로 어절들을 군집화시킨 후 발성을 하게 되며, 청자는 그러한 발화열에 대한 군집화의 경계 위치 정보를 화자의 의도를 이해하는 데 이용하게 된다. K-ToBI 는 한국어의 운율 정보를 표면적인 정보로 표기하는 방법 중의 하나로서, 여기에서는 이러한 발화열 중간에 위치하게 되는 끊김에 대한 정도를 Break Indices 층을 두어 별도로 정의하고 있다. 본 논문에서는 문서 음성 변환 시스템(Text-to-Speech system)의 자연성 향상에 큰 영향을 미치는 이러한 Break Indices 정보를 레벨의 정도에 따라 운율구 경계 유무와 경계 레벨로 정의하고 이를 예측하는 시스템에 있어 성능을 향상시키고자 한다. 기존의 운율 생성에 관한 연구에서는 CART를 이용하여 운율구의 경계 유무에 대한 정보를 예측하고 있으며, 영어권에서는 이 방법을 이용하여 억양구 경계의 위치를 비교적 높은 정확률로 예측하고 있다. 유럽권에서는 의존 구문트리를 이용하여 얻어진 구 정보를 운율구 경계로 근사 정의하는 연구가 보고되고 있다.

본 연구에서는 기존의 CART 모델에 근거하여 운율에 영향을 끼치는 언어 정보들 중에서 특징 변수들을 추출한 뒤, 이를 SVM 모델에 적용시켜 예측률의 향상을 유도한다. 일단, 각각의 문장에 대해 품사와 구문 정보를 포함하는 기본적인 언어 정보들을 기반으로 하여 총 26개의 특징 변수들을 제안한 후, 이 중 12개의 특징 변수만을 모델 생성에 적용한다. Break Indices 정보와 함께 구축한 코퍼스를 이용하여 해당 모델을 생성하고, 이에 대한 예측률을 CART 모델의 예측률과

비교함으로서 SVM 모델을 이용한 운율구 정보 예측이 보다 우수함을 보인다.

본 논문의 구성은, 우선 다음 장에서 K-ToBI 레이블링 시스템에 대해 간략히 언급하고 3장에서는 실험에 사용된 코퍼스와 통계 모델에 대하여 알아 본다. 4장에서는 코퍼스의 구축과 모델을 생성하는 과정에 대해 살펴 보고, 5장에서는 운율구 경계 및 레벨 정보의 예측에 대한 실험 결과를 다룬다. 그리고 마지막으로 6장에서 결론을 맺는 순서로 이루어져 있다.

2. K-ToBI 레이블링 시스템

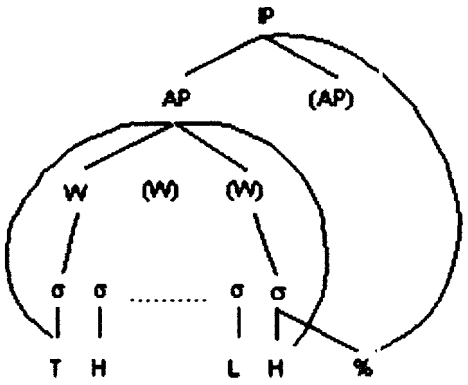

그림 1: K-ToBI의 구조

ToBI(Tones and Break Indices)는 1992년 제안된 레이블링 시스템으로, 다양한 실험이 ToBI 레이블링된 코퍼스 상에서 이루어져 오고 있다[4]. 시스템은 다중의 층으로 구성되어 있으며, 각 층은 발화열 내의 운율 정보를 표현한 기호를 포함하고 있다. K-ToBI 는 ToBI 로부터 고안된 한국어 운율 표기 규약으로, K-ToBI 시스템에서의 한국어 운율 구조는 그림 1과 같이 정의된다[4]. 즉, 발화열(U)는 한 개 또는 그 이상의 억양구(IP)로 이루어져 있고, 억양구는 한 개 또는 그 이상의 강세구(AP)와 경계 톤으로 이루어져 있다. K-ToBI 는 또한, 네 가지의 Break Index 값을 정의하고 있는데 이는 다음과 같다: 접어간의 명료한 운율의 영향인 경우에는 0; 접착 현상 등의 영향이 일어나지 않는 구 내부의 어절 경계의 경우에는 1; 주관적으로 강한 휴지를 느끼지 못하는 최소의 구 간의 분리 경계에 대해서는 2; 주관적으로 강한 휴지를 느끼는 구 간의 분리 경계에 대해서는 3의 값을 가진다. 이 중 2와 3은 각각 통상적으로 강세구 경계와 억양구 경계와 대응되는데, 여기서 Break Index 값은 단순히 외견상의 명백한 톤을 상징하는 접합이 아닌 분리도에 대한 레이블러의 주관적인 감각을 암시하고 있기 때문에, Break Indices 층의 정보는 Break Index 레벨 2와 3에 대한 톤 층의 정보와 완전하게 중복되지는 않는다. 그 밖에 톤이 Break와 어울리지 않는 경우 감지된 접합 정보 뒤에 포함시키는 'm' (eg. 2m, 3m) 이나 경계 레벨이 확실하지 않을 경우 포함시키는 '-' (eg. 1-, 2-) 등이 있지만 본 실험에서는 0~3 사이의 값만을 레벨 정보로 이용하였다.

3. 코퍼스 구성과 통계 모델

3.1 코퍼스 구성

본 연구에서 사용된 코퍼스의 정보는 다음과 같다. 여러 장르에서 발췌한 550 문장에 대해 기존의 구문분석기와 형태소 분석기를 이용하여 구문 분석 정보, 형태소 정보 등을 얻어낸 후 이에 대한 오류를 수정하였다[2]. 여기서 구문 분석 정보는 의존 트리의 형태로 표현되는데[5], 의존 트리는 문장을 구성하는 어절간의 관계를 지배어절과 의존어절로 표현하는 것으로서 모든 어절은 그에 대응되는 지배어절을 갖게 되고, 문장의 마지막 어절은 항상 자신을 지배어절로 갖게 된다. 또한, 영어의 경우와 달리 한국어의 경우는 지배어절이 항상 의존어절의 뒤에 위치하는 것을 발견할 수 있다[3]. 문장에 대한 형태소 분석 정보는 통상적으로 단일의 품사열 형태로 나타나지만 본 연구에서는 각 어절을 이루는 단어 성분에 대한 품사정보를 어절 내에서의 단어의 위치에 따라 좌품사 / 우품사로 구분하여 각각을 특징 변수에 포함시켰다. 이에 대한 한 예를 표 1에 보였다.

표 1: 운율구 코퍼스 구축을 위한 기본 정보

Index	어절	좌 / 우 품사	지배소 위치
0	요즘	ncnc	1
1	백화점에	ncpa	5
2	호남지역	ncnc	3
3	특산물전	ncnc	5
4	열풍이	ncps	5
5	불어닥치고	adad	6
6	있습니다	vjef	6

각 문장에 대한 Break Indices 정보는 음성 파형 분석기

의 결과를 참조해 가면서 청각적으로 발화의 끊김을 지각시키는 경계의 위치를 수작업으로 레이블링한 후, 특징 변수들과 함께 코퍼스에 포함시켰다. 운율구의 경계 위치는 일반적으로 억양의 급변화, 각 어절의 마지막 음절에 대한 장음화 그리고 휴지의 삽입과 같은 음성 신호에 있어서의 현상 등이 나타나게 된다. 본 연구에서는 K-ToBI 레이블링의 정의에 입각하여 문장을 이루는 어절 사이에서만 운율구 경계가 나타난다고 가정하였으며, 휴지가 발생되지 않는 구간에 대해서는 Break Indices 레벨 정보를 0으로 설정하였다. 그 외에, 문장과 문장의 사이를 제외한 나머지 어절간의 경계에 대해 그에 적합한 레벨 정보를 레이블링하였다. 코퍼스는 어절간의 경계만을 대상으로 해서 구축되었으며 각각의 경계 정보는 특징 변수 정보를 포함한다. 총 550 문장 중 어절간의 경계인 운율구 경계 후보의 개수는 10620 개였으며, 전체 코퍼스를 학습 코퍼스(전체 코퍼스의 70%, 7433개)와 테스트 코퍼스(전체 코퍼스의 30%, 3187개)로 나눈 후, 각각을 모델 생성과 테스트에 이용하였다. 또한 두 코퍼스의 통계치를 비교한 결과, 전체 코퍼스의 그것과 거의 비슷하였다.

3.2 Support Vector Machine

본 실험에서는 경계 레벨 정보를 예측하는 통계 모델에 있어 기존의 트리 기반 모델 대신 SVM을 이용한다. CART의 경우 기존의 통계적인 방법을 통한 경계 위치와 휴지기간 등의 운율구 정보 예측에 대한 연구에 주로 쓰이고 있는 모델이며, SVM은 최근 패턴 인식과 회귀 문제에 대한 해결 방법으로 널리 쓰이기 시작하고 있는 모델이다.

Support Vector Machine(SVM)은 Vapnik에 의해 제안된 통계적 학습 이론에 기반한 Universal Approximator로서, 기존의 학습 이론에서는 볼 수 없는 여러 특징들과 함께 뛰어난 empirical performance를 보이고 있어 많은 관심을 끌고 있다[7]. SVM이 가지는 특징 중의 하나는 새로운 데이터 샘플에 대해서 인식 오류율을 최소화하는 최적화된 분리경계면의 검색을 시도한다는 것이다. 여기서 SVM은 이론에 입각한 구조적인 위험 감소화(Structural Risk Minimization)를 이용하여 일반화 오류를 감소시키는 방법을 취하고 있으며, 입력 벡터들을 선형적인 분리경계면(linear hyperplane)이 얻어질 수 있는 높은 차원의 특징 공간(Feature Space)으로 맵핑을 시켜 주는 비선형 함수(non-linear function) 근사화 기능 역시 제공하고 있다. SVM은 분리경계면과 가장 가까운 거리에 위치하는 각 클래스의 데이터 포인트를 support vector라 정의하고, 분리경계면과 support vector와의 거리인 마진(margin)을 최대화하여 최적의 분리경계면(optimal hyperplane)을 찾아 낸다. 다음은 이를 식으로 나타낸 것이다.

$$f(x) = \sum_{i=1}^{l} y_i \alpha_i \bullet k(x, x_i) + b$$

$k(*,*)$ 은 커널 함수(kernel function)를 의미하며, $f(x)$ 는 x 의 레벨을 정의한다. 여기서 분리경계면을 찾는 과정은 $\alpha_i > 0$ 인 α_i 을 결정하는 과정이라고 볼 수 있으며, 조건을 만족하는 α_i에 부합하는 모든 벡터 x_i 는 분리경계면의 support vector로서 정의된다. SVM의 장점 중의 하나는 이러한 support vector의 개수를 적게 함으로서, 보다 컴팩트한 분류함수(classifier)를 생산하도록 한다는 것이다. 본 실험에서는 GNU버전의 S라고 할 수 있는 GNU/R 환경에서 SVM 라이브러리 패키지를 이용하여 SVM 모델을 생성하고 실험에 적용하였다.

4. 운율구 코퍼스 구축 및 모델 생성

4.1 최적의 특징변수 추출 및 코퍼스 구축

본 연구에서는 3.1절에서 언급하였던 기본 정보를 바탕으로 일반적으로 문음성 코퍼스에서 쓰여지는 26개의 특징변수들을 우선적으로 제안한다. 최적의 특징변수를 추출해내기 위한 방법으로 CART에서 트리를 확장할 경우 분리의 기준으로 사용되는 deviance를 이용한다. 우선 제안된 모든 특징 변수들을 대상으로 트리를 생성하고 과생성된 트리 모델을 대상으로 deviance가 가장 낮은 지점에서의 터미널 노드를 탐색한 후, 이 정보를 기반으로 pruning을 수행한다. 트리 모델 상에서 상위 노드일수록 데이터 분류시 불순도 양의 감소분을 최대화한다는 것에 근거를 두어 pruning된 트리에서 최상위 노드부터 터미널 노드까지 사용된 12개의 특징변수들을 최적의 특징변수들로 간주하고 이를 코퍼스로 구축하였다. 다음에 각 변수가 코퍼스 내에서 가지는 이름과 함께 변수들이 최종적인 예측 결과에 미치는 영향력을 간략하게 타진하여 보았다. 사용된 변수 이름에서 앞이 C인 변수는 실변수임을 나타내는 것이며, D인 변수는 카테고리 변수임을 나타내는 것이다.

Csylloc : 관측 어절에 대한 어절 내의 음절의 개수를 나타낸다. 이 값이 클 경우 현 관측 어절의 길이가 길다는 것을 의미하므로, 상대적으로 경계 레벨이 높게 측정될 가능성이 높다.

Cdlendirec : 관측 어절부터 지배어절까지의 음절 단위 거리를 나타낸다. 관측 어절과 지배어절간에 존재하는 음절의 수가 많다는 것은 곧 어절들이 이루고 있는 군집의 크기가 크다는 것을 의미하므로 관측 어절 바로 다음에 높은 경계 레벨이 올 가능성이 높다.

Ctotallen : 전체 문장을 구성하고 있는 음소의 개수를 나타낸다. 문장을 이루는 음소의 개수는 전체 문장에 대한 발화 길이와 비례한다. 따라서, 이 값이 작을 경우 발화 길이도 짧을 것이며 상대적으로 높은 경계 레벨의 발생 빈도도 작게 나타날 것이다. 또한, 전체 문장에 대한 정보인 관계로 문장 내 모든 경계 후보는 이 변수에 대해 같은 값을 가지게 된다.

Cwordloc : 전체 문장 내에서의 관측 어절의 위치를 나타낸다. 관측 어절이 문장의 앞 부분에 혹은 끝 부분에 가까울 경우, 경계 레벨이 낮게 측정될 가능성이 높다.

Dwordloc3 : 전체 문장 내에서의 관측 어절에 대한 상대적 위치를 3단계로 양자화한 값이다. 운율구 경계는 비슷한 크기의 단위로 발화가 이루어지는 것으로 기존 연구에서 보고된 바 있다. 이를 바탕으로 하여, 관측 어절의 상대적인 위치를 F/M/S 분류하여 사용하였다.

Drptag / Drstag : 관측 음소의 기준으로 전/후 음소에 대한 품사 정보. 경계 후보 바로 이전에 오는 어절의 마지막 음소가 가지는 형태소 정보와 바로 다음에 오는 어절의 가장 처음에 위치하는 음소가 가지는 형태소 정보를 의미한다.

Drightag : 관측 어절의 우품사 정보. 관측어절의 우측 품사의 경우 통상적으로 조사 내지는 어미가 되며, 이 중 주제격 조사가 사용되었을 때 운율구의 경계가 발생될 확률이 높다.

Dnltag : 관측 어절을 기준으로 다음 어절에 대한 좌품사 정보. 관측 어절의 바로 이후에 위치하게 되는 어절의 좌측 품사는 일반적으로 체언 내지는 용언이 오게 되는데, 이 중 용언이 사용되었을 경우, 낮은 경계 레벨이 발견될 가능성이 높다.

Ddirltag / Ddirrtag : 관측 어절에 대한 지배어절의 좌/우 품사 정보. 문장을 이루는 마지막 어절을 제외한 모든 어절들에 대해 지배어절은 항상 의존어절 뒤에 나타나므로, 지배어절 바로 다음에 높은 경계 레벨이 발생했을 경우 의존어절에서는 그와 같은 경계 레벨이 발생되지 않을 확률이 높다.

Ddirptag : 관측 어절의 지배어절을 기준으로 그 이전에 오는 어절의 우품사 정보. 일반적인 관측 어절과 마찬가지로 지배어절 이전의 어절 품사에 따라 지배어절 바로 뒤에 오는 경계 레벨이 그 영향을 받을 수 있으며, 이는 곧 의존어절인 관측어절 이후의 경계 레벨에 영향을 줄 수 있다.

4.2 모델 생성 및 튜닝

본 실험을 위해 GNU/R 환경에서의 SVM 라이브러리를 사용하여 SVM 모델을 학습시켰다. SVM의 파생 형태인 ν-SVM을 적용하여 모델을 생성하였으며[7], 커널 함수로는 Radial Basis 함수를 이용하였다[7]. ν-SVM은 다소 직관적이지 못한 기존 SVM에서의 C 조정 계수를 0과 1사이의 값을 가지는 ν로 대체한다. 일반적으로 모델을 생성하는 데에 있어 ν가 0에 가까울수록 Support Vector의 수는 감소하며, 1에 가까울수록 그 수가 증가한다. 표 2에서 알 수 있듯이 ν가 0.01을 가질 경우 각 코퍼스에서 가장 높은 예측률을 보이고 있다. 따라서 ν를 0.01로 가지는 모델을 최적의 모델로서 선정하였다.

표 2: ν에 따른 SV의 수와 예측률의 변화

	# of SV	학습 코퍼스	테스트 코퍼스
0.0001	600	32.23%	31.31%
0.001	3226	60.31%	52.30%
0.002	3947	77.66%	64.16%
0.004	4420	90.77%	73.64%
0.006	4589	98.77%	78.28%
0.008	4695	98.82%	78.69%
0.01	4762	99.94%	80.16%
0.02	4960	99.90%	79.69%
0.04	5171	99.58%	79.51%
0.06	5236	99.27%	79.29%
0.08	5291	98.88%	78.94%
0.1	5344	98.52%	78.60%

5. 운율구 경계 및 레벨 예측

네 단계의 경계 레벨 정보(0/1/2/3)를 갖는 학습코퍼스로 생성한 모델을 대상으로 테스트코퍼스를 사용하여 실험한 결과 표 3과 같은 예측률을 얻을 수 있었다. CART를 사용하여 생성한 결정 트리와 SVM을 사용하여 생성한 모델에 대한 confusion matrix를 함께 보였다.

표 3: CART 모델과 SVM 모델에 대한 경계레벨 예측 결과

(CART)

	S0	s1	S2	s3
s0	281	181	65	18
s1	118	605	187	59
s2	51	235	561	168
s3	33	71	195	359

Accuracy: 56.66%

(SVM)

	S0	S1	S2	s3
s0	482	64	16	10
s1	80	804	64	36
s2	40	105	734	75
s3	34	66	42	536

Accuracy: 80.16%

다음으로 (0/1)과 (2/3)의 레벨 정보에 대해 이를 각각 0, 1로 근사화하여 운율구의 경계에 대한 유무 정보로 가정하고 마찬가지로 학습코퍼스를 통하여 각 모델을 생성한 후 테스트코퍼스를 통하여 실험하였다. 이에 대한 confusion matrix는 다음과 같다.

표 4: CART 모델과 SVM 모델에 대한 경계유무 예측 결과

(CART)

	S0	S1
s0	1239	315
s1	369	1264

Accuracy: 78.53%

(SVM)

	s0	s1
s0	1373	165
s1	176	1473

Accuracy: 89.30%

표 3과 표 4에서 알 수 있듯이, 기존의 예측 모델인 CART와 비교하였을 때 본 방법은 경계의 레벨 예측과 유무 예측에서 각각 24%와 11%의 향상된 예측률을 보여주고 있다. 따라서, 문장에서 추출된 여러 언어 정보들을 토대로 운율구의 경계를 예측하는 방법에서 트리 기반 모델에 비해 본 논문에서 제시하고 있는 접근법이 보다 효과적인 접근법임을 입증하고 있다.

6. 결론

본 논문에서는 기존의 연구에서 사용한 트리 기반 모델을 대체하는 Support Vector Machine을 분류 함수로 적용하여 이를 통해 예측 시스템의 성능 향상을 유도하였다. K-ToBI 기반의 Break Indices 정보를 운율구의 경계 정보로서 레이블링하였고, 우선적으로 제안한 특징 변수들 중에서 트리 모델을 사용하여 최적의 특징변수들을 추출하였다. 구축한 코퍼스를 토대로 실험한 결과를 통하여 기존의 CART 방법에 비하여 SVM 모델이 운율구 예측을 위한 모델링에 보다 우수한 접근법이라는 것을 입증하였다. 또한, 기존의 향상된 트리 기반 모델을 이용한 운율구 경계 예측에 대한 정확률로서 보고되어 있는 85%에 비해 SVM을 이용한 모델은 약 5% 정도의 향상된 결과를 보였으며, 4단계로 이루어진 경계 정보에 대한 예측률도 80%대의 수치를 보였다.

향후 연구 방향으로는, 우선 SVM 모델을 위한 특징변수를 제시함에 있어 보다 분석적인 접근법이 요구된다. 그리고 K-ToBI를 근간으로 하는 운율의 구성성분에 대한 보다 체계적인 표기법이 꾸준히 연구되어야 할 것이며, 그에 따라 세부적으로 표면화된 운율 정보를 기반으로 하는 전체적인 운율 생성 모델의 연구와 예측 모델에 대한 연구가 병행되어야 할 것이다.

참고 문헌

[1] 이상호, 오영환, "CART를 이용한 운율구 추출 및 음소 지속 시간 모델링", 한국음향학회 학술발표대회 논문집 제 17권 1호, pp. 135-138, 1998

[2] 이상호, 오영환 서정연, "한국어 문서 음성 변환 시스템을 위한 문서 분석기", 한국음향학회지, Vol. 15, No.3, pp. 50-59, 1996

[3] 김창현, 김재훈, 서정연, "지배가능경로를 이용한 오른쪽우선 구문분석" 제 5회 한글 및 한국어 정보처리 학술발표 논문집, pp. 35-44, 1993

[4] M. Beckman and S.A. Jun. K-ToBI(Korean ToBI) Labeling Conventions, 1996.

[5] M.A. Covington, A dependency parser for variable word-order languages, *Research Report AI-1990-01*, Artificial Intelligence Programs, The University of Georgia, 1990.

[6] Y.J. Kim, H.J. Byeon, Y.H. Oh. "Statistical Prosodic-Boundary Prediction for Korean Text-to-Speech Convention", submitted to *Computer Processing of Oriental Languages*, 1999.

[7] Klaus-Robert Muller, Sebastian Mika, Gunnar Ratsch, Koji Tsuda, Bernhard Scholkopf. "An Introduction to Kernel-Based Learning Algorithms", *IEEE Transactions on Neural Networks*, Vol. 12, NO. 2, March, 2001

[8] J. Allen. "Overview of Text-to-Speech Systems". In S. Furui and M.M. Sondhi, editors, *Advances in Speech Siginal Processing*, pages 741-790. Marcel Dekker, 1991.

[9] M. Ostendorf, N. Veilluex. "A Hierarchical Stochastic Model for Automatic Prediction of Prosodic Boundary Location", *Computational Linguistics*, Vol 20, pages 27-54, 1994.

[10] J.Hirschberg and P.Prieto. "Training intonational phrasing rules automatically for English and Spanish text-to-speech". *Speech Communication*, 18(3):281-290, 1996.

[11] C.W. Wightman, S. Shattuch-Hunfnagel, M. Ostendorf and P.J. Price, "Segmental durations in the vicinity of prosodic phrase boundaries," *J.Acoust. Soc. Am.*, Vol. 91, No.3, pp. 1707-1717, 1992.