

합성음성 경보의 주관적 위급도에 관한 연구

박 경수 · 장 필식

한국과학기술원 산업공학과

ABSTRACT

This paper presents an experimental study of the relationship between sound parameters of synthesized voice warning and perceived(psychoacoustic) urgency. Eighteen subjects participated in two experimental sessions to evaluate and quantify the effects of the voice parameters. Experiments showed that speech rate, fundamental frequency and voice types have clear and consistent effect on perceived urgency. The results of these experiments can be applied to the improvement of existing auditory warning systems and the design of new systems.

1. 서론

청각적 표시 장치는 청각 계통의 고유한 특성 때문에 다른 感官(sensory modality)을 이용하는 표시 장치들과 비교해서, 특정한 상황에 있어 여러 가지 장점을 가진다. 이 때문에 청각적 표시 장치는 위급한 상황에 대한 정보를 전달하는 기능을 가지는 경계 및 경보 신호를 위한 용도로 많이 사용된다. 하지만 기존의 경계 및 경보 기구들, 즉 부저(buzzer), 경적(horn), 호각(whistle), 사이렌(siren), 전종(bell) 등에 의한 경보에 대해서 최근, 여러 가지 문제점들이 지적되고 있다. 이러한 단점들에 대처하고, 기존의 청각적 경보를 개선시키거나 새로운 청각적 경계, 경보를 설계하기 위한 접근 방법으로는, 합성 음성(synthesized speech)을 경보 체계에 이용하는 방법과 디지털 방식을 이용하여 새로운 경보음을 만들고, 여기에 주관적 위급도(perceived urgency)와 같은 부가적인 정보를 포함시키는 방법의 두 가지로 대별할 수 있다.

합성 음성은 합성 음성 자체로써 경보의 기능을 가질 수 있으며, 기존의 청각적 경보 체계와는 달리 학습 없이 인간에게 친숙한 형태로 정보를 전달할 수 있다는 장점 때문에 최근, 새로운 청각적 경보의 형태로써 많은 연구들을 통해 활용 범위가 넓어지고 있다. 하지만 인간의 음성과 너무 비슷한 경우, 경보가 제대로 수신자에게 전달되지 않는다거나, 통신을 방해하게 되기도 하며, 체계 성능에 대한 필요 이상의 과신이 생길 수 있다는 단점이 있다.

디지털 방식을 이용하여 새로운 경고 음을 만들고, 여기에 주관적 위급도와 같은 정보를 포함시키는

경우, 이러한 정보는 인간이 직관적으로 느낄 수 있는 정보라는 면에서 장점을 가지며, 이러한 장점 때문에 경보음의 수가 많지 않을 경우에는 청각적 경보 체계의 성능을 크게 향상시킬 수 있다고 알려져 있다. 하지만 이렇게 경보음에 위급도를 암호화시키더라도 경보음 본래의 특성상, 기존의 경보 장치와 마찬가지로, 경보음의 수가 많아 생기게 되는 문제점에 대한 기본적인 해결책은 되지 못할 것으로 생각된다.

2. 합성 음성 경보에의 주관적 위급도 암호화

본 연구에서는 위에서 언급한 단점들을 보완하여 합성 음성 체계의 성능을 향상시키는 인간공학적 접근 방법의 하나로써, 합성 음성 경보에 인간의 주관적 위급도를 포함시키는 방법에 대하여 논해 보고자 한다. 즉 합성 음성의 음향학적인 파라미터를 포함한 여러 가지 요소들을 합성 음성의 이해도(intelligibility)를 저하시키지 않는 범위 내에서 변화시켜, 상황의 위급도와 이에 상응하는 인간의 주관적 위급도를 합성 음성 경보음에 암호화하는 것이다. 이렇게 위급도를 암호화한 합성 음성 경보는 '정보 전달'과 '경보'의 기능을 가지는 합성 음성 자체의 장점을 가지게 될 것이며, 인간의 음성과 구별되는 경보음으로써의 장점을 가지게 되므로, 합성 음성 경보 체계의 성능을 향상시킬 수 있으리라 생각된다.

이를 위해서 본 연구에서는 인간의 주관적 위급도에 영향을 미치는 합성 음성의 요소들을 실험을 통해 찾아내고, 이러한 요소들이 어느 정도 주관적 위급도에 영향을 미치는 지를 평가하고자 하며, 경보가 발생하는 상황의 위급한 정도와 주관적으로 느끼는 위급도와의 관계를 정량화 하고자 한다. 본 연구에서 인간의 주관적 위급도에 영향을 미치는 합성 음성 요소들을 찾아내기 위해 고려한 요소들로는 합성 음성 경보에 사용되는 어휘(vocabulary), 경보의 길이(vocabulary size), 기초 진동수(Fundamental Frequency, Pitch), 음성 경보의 빠르기(speech rate), 합성 음성 경보의 음색 등이 있다.

이러한 요소들 중 어느 요인이 인간의 주관적 위급도에 영향을 미치며, 어느 정도 영향을 미치는 지를 평가하기 위해서 본 연구에서는 Magnitude Estimation 기법을 사용하게 되며, 외부 자극인 상황의 위급도와 인간이 주관적으로 느끼는 주관적 위급도의 관계는 다음과 같은 Stevens의 Power Law를 사용하였다. 또한 실험후 측정된 자료들에 대한 타당성을 검증하기 위하여 Cross-Modality Matching Paradigm을 이용하였다.

$$S = kO^m \quad \left\{ \begin{array}{l} S : \text{주관적인 지각 강도(Subjective Magnitude)} \\ O : \text{외부 자극 강도(Stimulus Magnitude)} \\ k : \text{비례상수} \\ m : \text{지수(Exponent)} \end{array} \right.$$

3. 실험

3.1 피실험자

실험에 참여한 총 인원은 18 명으로, 대부분 대학원에 재학 중인 23세에서 29세까지의 학생들로 구성되었다. 피실험자의 성별은 남자가 15명, 여자가 3명이었으며, 모든 피실험자가 개인용 컴퓨터를 다루어 본 경험이 있었고, 실험전 청각의 이상 유무를 확인하였다.

3.2 실험 장치

실험에 사용되는 합성 음성 경보는 개인용 컴퓨터에서 사용 가능한 sound board(Sound Blaster AWE32)와 이 board를 사용하는 음성합성 소프트웨어인 Text Assist(Creative Labs)를 사용해 제작하였다. 합성 음성은 같은 사양의 sound board를 장착한 두 대의 PC를 연결한 뒤, sampling rate 20kHz, 16bit WAV file format 형태로 저장하여 실험에 사용하였다. sound board를 제어하고, 실험을 진행시키는 프로그램은 C 언어로 구현하였으며, 주위의 잡음을 차단하기 위하여 sound board에 헤드폰(Philips SBC 3375)을 연결하여 사용하였다. 기초 진동수는 16비트 20kHz sampling rate 형태로 저장된 합성 음성을 8비트 10 kHz sampling rate 형태로 변환하여 SIFT(Simplified Inverse Filter Tracking) algorithm을 사용하여 측정하였다.

3.3 실험 절차

실험 절차는 우선 피실험자에게 Magnitude Estimation 기법을 이해시키고, 피실험자의 비율 판단 능력을 검증하며, 주관적 위급도를 평가할 때 사용하는 수치 평가와 선을 긋는 경향을 파악하여 본 실험 결과를 보정하기 위해서 예비 실험을 실시하였다. 그 다음 이러한 예비 실험에 합격된 피실험자들을 대상으로 하여 실험과 결과 분석을 행한다. 본 실험에서 실험 변수로 사용되는 합성 음성 요소들과 각 변수의 수준은 다음 표와 같으며, 각 실험 조건을 2회 반복하여 총 24회의 실험이 이루어졌다.

표 1. 실험 변수

실험 변수(factor)	변수의 수준(level)
합성 음성의 음색	2
기초 진동수(fundamental frequency)	2
합성 음성 경보의 빠르기(speech rate)	3

자극으로 주어지는 합성음성 경보에 대한 피실험자의 주관적 위급도는 한 자극에 대하여 선의 길이 (Line Production: LP)와 수치(Numerical Estimation: NE) 두 가지로 평가되는데, 이 두 가지 평가치를 통합하여 실험의 종속변수로 사용하게 된다. 각 실험 조건에서 얻게 되는 이 두개의 평가치에 대해서는 Regression Bias를 교정하기 위하여 Lodge가 제안한 다음과 같은 식을 이용하여 통합하였다.

$$IM = (L^{1/n1} \cdot N^{1/n2})^{0.5}$$

- IM : 통합된 측정치(Integrated Measure)
- L : 선의 길이(Line Production)
- N : 수치(Numerical Estimation)
- $n1$: 각 피실험자의 선을 긋는 경향에 관한 회귀식 계수
- $n2$: 각 피실험자의 수치 평가 경향에 관한 회귀식 계수

3.4 실험 결과

예비 실험에 참여한 피실험자 18명중, 비율 판단 능력에 대한 검증(예비 실험)에 합격한 15명을 대상으로 실험 결과 분석을 행하였다. 기초 진동수와 Speech Rate, 음색의 종류에 따라 피실험자들의 주관적 위급도가 영향을 받는지를 파악하기 위하여 분산 분석을 실시하였다. 표3에서 보는 바와 같이, 합성 음성의 음색의 종류는 5% 유의 수준에서 유의함을 보이고 있으며, 기본 주파수와 Speech Rate는 1% 유의 수준에서도 유의함을 보이고 있다. 즉 이 세 가지 요인들이 모두 주관적 위급도에 유의한 영향을 미치고 있음을 알 수 있으며, 분산 분석표에서 보는 바와 같이 이들 인자들간에 2인자 이상 교호작용은 통계적으로 유의하지 않는 것으로 판명되었다.

표 2. 전체 Model에 대한 분산분석 표

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	25	141370.0102	5654.8004	22.98	0.0001
Error	334	82205.9804	246.1257		
Corrected Total	359	223575.9906			
	R-Square	C.V.	Root MSE	URGENCY Mean	
	0.632313	41.08502	15.68839	38.1851929	

표 3. 각 독립변수들에 대한 분산분석 표

Source	DF	Anova SS	Mean Square	F Value	Pr > F
음성의 유형	1	1409.28258	1409.28258	5.73	0.0173 *
Speech Rate	2	97566.77730	48783.38865	198.21	0.0001 *
기본 주파수	1	6745.27582	6745.27582	27.41	0.0001 *
음성유형 * Speech Rate	2	6.09945	3.04972	0.01	0.9877
음성유형 * 기본 주파수	1	548.22002	548.22002	2.23	0.1365
Speech Rate * 기본 주파수	2	643.67954	321.83977	1.31	0.2718

VTYPE * SRATE * PITCH	2	126.30091	63.15045	0.26	0.7738
SUBJECT	14	34324.37461	2451.74104	9.96	0.0001

* 유의 수준 $\alpha=0.05$ 에서 유의

4. 결론 및 연구 방향

본 연구에서는 인간의 주관적 위급도에 영향을 미치는 합성음성의 요소들을 실험을 통해 찾아내고, 이러한 요소들의 영향을 평가하였다. 실험 결과 음성의 유형(음색), speech rate, 기본주파수(fundamental frequency) 등의 요인들이 주관적 위급도에 유의한 영향을 미치는 것으로 나타났으며, 이들간의 2인자 이상 교호 작용은 없는 것으로 평가되었다. 이러한 결과를 바탕으로 하여, 각 요인들에 대해 Stevens의 Power Law를 이용, 외부 자극인 상황의 위급도와 인간이 주관적으로 느끼는 주관적 위급도의 관계를 정량화 하는 작업이 수행될 것이다. 이를 통해 음성의 이해도(intelligibility)를 저하시키지 않는 범위 내에서 요인들을 변화시켜, 상황의 위급도와 이에 상응하는 인간의 주관적 위급도를 합성 음성 경보음에 암호화하고자 한다. 이렇게 위급도를 암호화한 합성음성 경보는 '정보 전달'과 '경보'의 기능을 가지는 합성음성 자체의 장점을 가지게 될 것이며, 인간의 음성과 구별되는 경보음으로써의 장점을 가지게 되므로, 합성음성 경보 체계의 성능을 향상시킬 수 있으리라 생각된다.

참 고 문 헌

- [1] Carl E. Williams., Stevens. 1972 'Emotion and Speech: Some Acoustical Correlates 'J.A. Acoust. Soc. Am Vol 52 Number 4 part 2, 1238-1250.
- [2] Edworthy, J., Loxly, S. and Dennis, I. 1991 'Improving auditory warning design: relationship between warning sound parameters and perceived urgency' Human Factors33(2), 205-231.
- [3] Hellier, E , Edworthy, J. and Dennis, I. 1993 'Improving auditory warning design : quantifying and predicting the effects of different warning parameters on perceived urgency' Hum Factors.35(4), 693-706.
- [4] Markel, J.D. et & Gray. 1976 'Linear Prediction of Speech' Springer-Verlag.
- [5] Kerr, J.H. 1985 'Auditory warnings in intensive care units and operating theatres' Ergonomics International, 85, 172-174.
- [6] Lodge, M. 1981 'Magnitude Scaling: Quantitative measurement of opinions' Beverly Hills, CA: Sage.
- [7] Momtahan, K.L., Hetu, R. and Tansley, B.W. 1993 'Audibility and identification of auditory alarms in operating rooms and an intensive care unit' Ergonomics 36(10), 1159-1176.

- [8] Rabiner, L. R., M. J. Cheng, A. E. Rosenberg, and C. A. McGonegal. 1976 'A Comparative Performance Study of Several Pitch Detection Algorithms' IEEE Tras. Acoust., Speech, and Signal Process., vol ASSP-24, 399-417.
- [9] Sorkin, R., Kantowitz, B., and Kantowitz, S. 1988 'Likelihood alarm displays' Human Factors, 30, 455-459.
- [10] Stevens S. 1957 'On the psychophysical law' Psychological Review, 64, 153-181.
- [11] 박경수. 1980 '인간공학', 영지문화사.