

화자의도예측 파라미터를 이용한 조타명령 음성인식 시스템의 개선

문 성 배[†]

(원고접수일 : 2008년 5월 2일, 원고수정일 : 2008년 7월 16일, 심사완료일 : 2008년 7월 18일)

Enhancement of Ship's Wheel Order Recognition System using Speaker's Intention Predictive Parameters

Serng-Bae Moon[†]

Abstract : The officer of the deck(OOD) may sometimes have to carry out lookout as well as handling of auto pilot without a quartermaster at sea. The purpose of this paper is to develop the ship's auto pilot control module using speech recognition in order to reduce the potential risk of one man bridge system. The feature parameters predicting the OOD's intention was extracted from the sample wheel orders written in SMCP(IMO Standard Marine Communication Phrases). We designed a pre-recognition procedure which could make some candidate words using DTW(Dynamic Time Warping) algorithm, a post-recognition procedure which made a final decision from the candidate words using the feature parameters. To evaluate the effectiveness of these procedures, the experiment was conducted with 500 wheel orders.

Key words : Auto pilot(자동조타장치), Speech recognition(음성인식), Feature parameter(특정 파라미터), Wheel order(조타명령), Speaker intention(화자의도), SMCP(표준해사통신용어), DTW(동적시간신축)

1. 서 론

2000년 국제해상인명안전협약(SOLAS협약) 제 5장이 전면 개정되고 전자해도표시·정보시스템(ECDIS), 선박자동식별장치(AIS), 항해자료기록기(VDR) 및 음향수신기(Sound Reception System)와 같은 IT융합 장비가 선박에 탑재되면서 해기사의 선교 근무환경 개선과 운항 효율화를 도모할 수 있게 되었다. 또한 선박자동화를 통하여 기관실의 제어가 가능해짐으로써 선교 1인 당직체

제(One man bridge system)의 기틀이 마련되었다^[1].

이와 같은 선박의 자동화뿐만 아니라 해운산업의 경영합리화 등은 선박 승무원 수를 감소시키는 계기가 되었다. 그러나 교통량이 많은 해역을 항해할 때 항해사가 단독으로 선박조선 업무를 수행하는 상황이 자주 발생하고 있어 돌발적인 위급상황에서 해양사고 발생 위험성이 그 만큼 커지고 있는 실정이다^[2].

최근 이러한 선교 1인 당직체제에서 조타수의 역

[†] 교신저자(한국해양대학교 항해시스템공학부, E-mail:msbae@hhu.ac.kr, Tel: 051)410-4280)

할을 대신할 수 있는 시스템으로 항해사의 음성 조타명령(Wheel order)만으로 자동조타장치(Auto pilot)를 제어할 수 있는 음성인식(Speech recognition) 기반의 자동조타장치의 개발에 관한 연구가 진행되고 있다^[1].

조타명령은 선박 충돌사고와 직결될 수 있기 때문에 높은 인식성능은 개발에 있어서 매우 중요한 요소이다. 그러나 지금까지 연구 또는 개발된 시스템은 선박의 소음환경, 독특한 조타명령의 발성스타일 및 음운특성을 고려하기 보다는 일반적인 음향학적 기법의 적용을 기반으로 하였기 때문에 실제 적용하기에는 낮은 인식률과 안전성의 문제로 크게 활용되지 못하고 있다^{[3],[4]}.

이 논문은 단순히 음향학적 인식기법의 개선을 통하여 인식성능을 높이기보다는 열악한 조타명령의 음성인식환경을 고려하여 항해사의 조타명령 발성의도를 예측할 수 있는 화자의도예측 파라미터를 패턴인식과정에 적용하여 인식률을 향상시킬 수 있는 후처리 음성인식 프로세스를 개발하고자 하는 것이다.

2. 화자의도예측 파라미터

음성이해(Speech understanding)는 음향학적 음성인식기법의 개선 이외에 화자의 발성 의도를 예측할 수 있는 지식을 음성신호의 음향학적 해독이나 구문처리 과정에 적용하여 화자가 원하는 것을 수행하도록 하는 것이다. 최근 이 기법을 도입하여 음성인식의 오류를 줄이려는 연구가 활발히 진행되고 있다^{[5],[6]}.

화물선 항해사관이 선박조종에 사용하는 조타명령문은 IMO SMCP(IMO Standard Marine Communication Phrases)에 제시되어 있다. Table 1은 SMCP에 제시된 총 23개 형식의 조타명령 예문을 나타낸 것이다.

23개 조타명령문 형식 중에서 "Steady", "Steady as she goes", "Finished with wheel", "Report if she dose not answer wheel", "Keep buoy on port side"은 음성인식을 통하여 정확하게 패턴이 결정되더라도 결국 동

작의 수행은 인간의 지적인 판단이 요구되는 것이라 할 수 있다. 따라서 이 연구에서는 상기 5가지 조타명령문을 인식대상 구문 종류에서 제외하여 총 18가지 형식의 조타명령문을 음성인식의 범위로 하였다. 그리고 이 18가지 조타명령문은 20개 단어의 조합을 통하여 구성될 수 있기 때문에 음성인식 대상 단어는 20개로 제한하였다.

한편 Table 1에서 'Ease-to'는 'Ease'와 'To'의 두 고립단어로 각각 인식되어야 하지만 조타명령으로 발성 시 휴지시간이 매우 짧고 'To'가 많이 약화되기 때문에 이 연구에서는 하나의 단어로 취급하여 인식 대상인 단어의 수는 19개로 처리하였다.

인식대상으로 선정한 조타명령문은 크게 타각(Rudder angle)을 제어하는 타각제어명령문과 선박의 침로를 지정하는 침로지정명령문으로 구분할 수 있다. 또한 명령문들을 구성하는 단어의 수에 따라서 1단어 명령문, 2단어 명령문, 3단어 명령문 및 4단어 명령문으로 분류할 수 있다.

Table 1 Standard wheel orders

Wheel Orders
Midships
Port five
Port ten
Port fifteen
Port twenty
Port Twenty five
Hard port
Starboard five
Starboard ten
Starboard fifteen
Starboard twenty
Starboard twenty five
Hard Starboard
Ease-to five
Ease-to ten
Ease-to fifteen
Ease-to twenty
Steady
Steady as she goes
Keep buoy/mark/beacon/... on port side.
Report if she does not answer wheel.
Finished with wheel.
Steer one two five

이 연구는 조타명령문을 구성하는 단어의 수와 위치(순서)와 같은 특징 파라미터를 바탕으로 화자의 의도를 예측함으로써 음성인식의 정확도를 향상시키기 위한 것이다. Table 2는 조타명령문을 구성하는 단어의 수에 따라 각 단어별로 발생될 가능성의 유무를 파라미터 값으로 나타낸 것이다. 그리고 Table 3은 명령문의 단어 순서에 따라 각 단어별로 발생될 가능성의 유무를 값으로 부여하여 나타낸 화자의도예측 파라미터의 값을 나타낸 것이다^[7].

3. 화자의도예측 음성인식 시스템

이 연구에서 개발하고자 하는 화자의도예측 기반의 조타명령 음성인식 시스템은 전형적인 음성인식 시스템을 바탕으로 기본적인 음성인식 과정을 수행한 후 후보 단어들을 먼저 결정하게 된다. 그 후 Table 2와 Table 3의 파라미터 값을 이용하여 화자의 의도를 추론할 수 있는 프로시저를 통하여 최종적인 인식 단어를 결정하도록 설계하였다. Fig. 1은 이 시스템의 전체적인 구성 및 흐름을 나타낸 것이다.

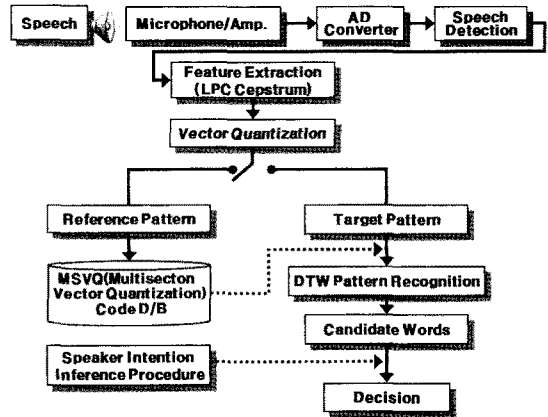


Fig. 1 Wheel order recognition system using speaker intention inference procedure

Table 2 Parameters depending on word number

Word number	1	2	3	4
Word				
One	0	0	0	1
Two	0	0	0	1
Three	0	0	0	1
Four	0	0	0	1
Five	0	1	1	1
Six	0	0	0	1
Seven	0	0	0	1
Eight	0	0	0	1
Nine	0	0	0	1
Ten	0	1	0	0
Zero	0	0	0	1
Fifteen	0	1	0	0
Twenty	0	1	1	0
Starboard	0	1	1	0
Port	0	1	1	0
Hard	0	1	0	0
Midships	1	0	0	0
Steer	0	0	0	1
Ease-to	0	1	0	0

Table 3 Parameters depending on word position

Word	First word	Second word	Third word	Fourth word
One	0	1	1	1
Two	0	1	1	1
Three	0	1	1	1
Four	0	0	1	1
Five	0	1	1	1
Six	0	0	1	1
Seven	0	0	1	1
Eight	0	0	1	1
Nine	0	0	1	1
Ten	0	1	0	0
Zero	0	1	1	1
Fifteen	0	1	0	0
Twenty	0	1	0	0
Starboard	1	1	0	0
Port	1	1	0	0
Hard	1	0	0	0
Midships	1	0	0	0
Steer	1	0	0	0
Ease-to	1	0	0	0

3.1 음성신호의 분석

먼저 조타명령 음성은 범용의 개인용 컴퓨터용 헤드셋에 장착된 콘덴서형 마이크로폰으로 수신되고, LM358 OP Amp. 소자를 기본으로 구성된 증폭기를 통하여 약 38dB 정도 증폭하였다. 그리고 이 신호는 Table 4와 같은 사양의 AD변환기를

통하여 디지털 신호로 변환되도록 시스템을 설계하였다.

Table 4 Specifications of AD converter

Maker	TiePie Engineering, Netherlands
Model	Handyscope 3
Interface	USB 2.0 & USB 1.1
Channels	AD : 2, DA :1
Resolution	12bits
Sampling Frequency	8kHz
Sampling Period /Sample Number	3.75sec/30,000samples

조타명령은 한 명령문을 구성하는 단어의 수가 적기 때문에 고립단어인식 기술을 적용하였다. 디지털 값으로 변환된 전체 신호데이터에서 음성신호는 유성음과 무성음의 식별과 고음질의 신호에서 묵음과 무성음의 식별이 용이한 단구간 에너지 함수법을 이용하여 탐지하도록 프로시저를 구축하였다[8].

음성인식에 필요한 특징 벡터열을 추출하는 기법에는 LPC(Linear Predictive Coding) 캡스트럼, PLP(Perceptual Linear Prediction) 캡스트럼, MFCC(Mel Frequency Cepstral Coefficient), Filter Bank Energy 등이 있다. 이 연구에서는 화자의 의도를 예측할 수 있는 파라미터의 적용을 통한 음성인식 시스템의 개선을 목적으로 화자종속방식을 기반으로 시스템을 설계하였기 때문에 음성발생기관의 모델에 근거를 두고 화자종속 음성인식 기법에 많이 사용되는 LPC 캡스트럼을 선정하였다.

이 음성인식 시스템에서는 Levinson-Durbin 알고리즘을 이용하여 LPC 계수를 연산한 후 LPC 캡스트럼 계수를 연산하도록 프로그래밍 하였다. 여기서 특징 벡터열인 LPC 캡스트럼 계수 c_m 은 (1.a)와 (1.b)로 나타낼 수 있다^[9].

$$c_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m} \right) c_k a_{m-k}, \quad 1 \leq m \leq p \quad (1.a)$$

$$c_m = \sum_{k=m-p}^{m-1} \left(\frac{k}{m} \right) c_k a_{m-k}, \quad p < m \quad (1.b)$$

단, a_m 은 LPC계수, p 는 LPC 차수, m 은 캡스트럼 차수

3.2 전처리 음성인식 프로세스

음성인식이라는 것은 비교의 기준이 되는 표준패턴과 입력패턴과의 유사도를 측정하는 것이라 할 수 있다. 여기서 이 패턴은 일반적으로 각 단어에서 추출된 특징 벡터열을 양자화하여 작성되는 코드북(Code book)에 해당하는 것이다^[10]. 이 코드북은 음성인식을 위한 시계열 음성데이터 단위별로 구성되는 코드북 숫자에 따라 단일섹션 코드북(Single section code book)작성법과 멀티섹션 코드북(Multi-section code book) 작성법으로 나눌 수 있다. 단일섹션 코드북으로 작성될 경우 인식률이 떨어지기 때문에 이 연구에서는 멀티섹션 코드북 작성법으로 표준패턴을 구축하였다^[11]. 멀티섹션 코드북의 경우 섹션의 사이즈는 100샘플, 중첩 사이즈는 50샘플로 설정하였다.

한편 벡터열을 양자화하는 기법으로는 K-means, LBG, LVQ 등이 있는데 이 연구에서는 가장 간단한 자율학습(Unsupervised learning) 알고리즘의 하나인 K-means를 채택하여 시스템을 프로그래밍 하였다.

3.3 후처리 음성인식 프로세스

전처리 프로세스의 벡터 양자화를 통하여 추출한 표준패턴과 입력패턴의 유클리드 거리(Euclidean distance)를 측정하여 간단하게 인식여부를 결정하는 방법도 있다. 그러나 이 패턴들에는 시간 축상에서의 변화정보를 가지고 있지 않기 때문에 정확한 패턴비교에 적용하기에는 어려움이 있다^[12]. 이러한 어려움을 극복하고 인식률을 향상시키기 위하여 이 시스템에서는 입력패턴(인식대상 패턴)과 표준패턴 사이의 시간적 차이를 DTW(Dynamic Time Warping) 기법을 이용하여 제거하도록 하였다.

DTW를 통하여 생성된 DTW 표준패턴과 입력 패턴 사이의 유클리드 거리를 연산하여 후보단어 (Candidate words)를 도출하고, Table 2와 Table 3에서 결정된 화자의도예측 파라미터 값을 후보단어에 곱하여 가장 유사도가 높은 단어를 최종의 인식된 어휘로 결정하는 알고리즘을 개발하였다.

먼저 음성신호 분석단계에서 조타명령문을 구성하는 단어의 수를 $wn(1 \leq wn \leq 4)$ 이라 하고 DTW 템플릿 매칭을 통하여 도출한 후보단어의 유클리드 거리를 D_{CW} 라 할 때, wn 파라미터 적용을 통하여 연산된 새로운 후보단어의 유클리드 거리 D_{NCW} 는 다음과 같이 나타낼 수 있다.

$$D_{NCW}(n, k) = D_{CW}(n, k) \cdot W_N(n, k) \quad (2)$$

단, W_N 은 단어 수에 의한 화자의도 예측 하중 값, k 는 인식대상 단어($1 \leq k \leq 19$),

$$1 \leq n \leq wn$$

그리고 (2)에서 새롭게 연산된 D_{NCW} 에 Table 3의 파라미터 값을 적용하여 후보단어들의 최종 유클리드 거리 D_{FCW} 를 다음 (3)을 이용하여 결정할 수 있다.

$$D_{FCW}(n, k) = D_{NCW}(n, k) \cdot W_P(n, k) \quad (3)$$

단, W_P 는 단어위치에 따른 화자의도 예측 하중 값,

$$1 \leq k \leq 19, 1 \leq n \leq wn$$

(3)의 연산과정을 통하여 도출된 후보단어들 중에서 유클리드 거리가 가장 작은 즉 유사도가 가장 높은 단어를 인식어휘 D_W 로 간주하도록 후처리 음성인식 프로세스가 설계되었는데, (2)와 (3)이 포함된 이 프로세스는 (4)와 같이 나타낼 수 있다.

$$D_W(n) = \min [D_{CW}(n, k) \cdot W_N(n, k) \cdot W_P(n, k)] \quad (4)$$

4. 실험방법 및 결과

4.1 실험방법

Table 5는 이 연구에서 구현한 화자의도예측 파

라미터를 이용한 후처리 음성인식 프로세스의 음성 인식 성능 개선 정도를 검증하기 위하여 선택한 기초 실험용 조타명령 구문의 종류를 나타낸 것이다. 실험용 조타명령 구문에서 타각지정 명령은 SMCP의 내용을 그대로 사용하도록 하였고, 침로 지정 명령문은 임의로 3개를 선정하여 총 20개 명령문으로 실험을 수행하였다.

그리고 음성인식용 어휘의 표준패턴을 작성하기 위한 훈련용 음성 데이터베이스와 조타명령 구문의 발성 실험자로 39세의 남성 1인을 선정하였다. 그리고 Table 5의 실험용 조타명령 구문은 랜덤하게 각 구문이 25회씩 발성되도록 하였다.

발성된 조타명령 구문들을 Visual C++ 6.0으로 작성된 음성신호의 검출, 전처리 음성인식 프로세스 및 후처리 음성인식 프로세스 프로그램을 통하여 처리되도록 하였다.

Table 5 Wheel orders for experiments

Type	Wheel orders	Word number	
Rudder angle	Midships	1	
	Port five	2	
	Port ten	2	
	Port fifteen	2	
	Port twenty	2	
	Hard port	2	
	Starboard five	2	
	Starboard ten	2	
	Starboard fifteen	2	
	Starboard twenty	2	
	Hard Starboard	2	
	Ease-to five	2	
	Ease-to ten	2	
	Ease-to fifteen	2	
	Ease-to twenty	2	
	Port twenty five	3	
	Starboard twenty five	3	
	Ship's Heading	Steer zero seven zero	4
		Steer one two five	4
		Steer two four nine	4

4.2 실험결과

화자의 의도를 추론하여 음성인식을 수행하는 후처리 음성인식 모듈의 인식성능을 검증하기 위하여

DTW기법만을 이용한 인식률과 화자의도예측 파라미터의 적용을 통한 인식률을 비교하는 실험을 수행하였고 그 결과는 Table 6과 같다. 이 연구에서의 음성 인식률은 조타명령 구문을 구성하는 각 단어별로 인식률을 산정한 것이 아니라 명령문을 구성하는 모든 단어를 모두 정확하게 인식하는 경우만 성공한 인식으로 간주하여 실험을 수행하였다. 따라서 이 시스템에서의 인식률은 단어별 인식률에 비해 다소 떨어진다고 할 수 있다.

DTW 패턴인식만을 적용한 21개 조타명령 구문의 전체 평균 인식률은 73.4%로 나타났지만 화자의 의도 추론용 특징 파라미터를 이용한 후처리 인식모듈의 경우는 88.8%로 나타나 15.4% 개선되는 것을 알 있었다.

Table 6 Result of wheel order experiments

Wheel orders	Dynamic time warping	Speaker's intention predictive parameters
Midships	100%	100%
Port five	80%	100%
Port ten	80%	100%
Port fifteen	64%	72%
Port twenty	64%	100%
Port twenty five	28%	96%
Hard port	84%	100%
Starboard five	84%	84%
Starboard ten	48%	52%
Starboard fifteen	80%	92%
Starboard twenty	84%	92%
Starboard twenty five	88%	96%
Hard Starboard	96%	96%
Ease-to five	60%	96%
Ease-to ten	24%	44%
Ease-to fifteen	68%	84%
Ease-to twenty	72%	92%
Steer zero seven zero	80%	88%
Steer one two five	100%	100%
Steer two four nine	84%	92%
average	73.4%	88.8%

특히 'Port twenty five'의 경우 DTW 패턴인

식에서는 28%의 인식률을 보였지만 화자의도 추론 인식과정에서는 96%로 나타나 인식률이 68% 개선되었다. 또한 'Port five', 'Port ten', 'Hard port'의 경우는 DTW 패턴인식에서는 90%이하의 인식률을 나타내었지만 화자의도 추론 인식과정에서는 100%의 인식 성능을 보였다. 한편 화자의도 추론 음성인식 결과에서 인식률이 80%이하인 'Port fifteen', 'Starboard ten', 'Ease-to ten'의 경우를 제외한 평균 인식률은 약 94.6%로 나타났다.

따라서 DTW 패턴인식과 같은 전처리 음성인식 프로세스에서 인식률 향상을 도모할 수 있는 특징 벡터 추출법 및 패턴인식 기법 등을 보완한다면 신뢰성 있는 조타기 음성제어 시스템의 실용화가 가능할 것으로 사료된다.

5. 결 론

이 연구에서는 높은 신뢰성이 요구되는 조타기의 음성인식 제어시스템을 개발하기 위하여 특징벡터 추출법 또는 패턴인식 기법의 적용을 통하여 인식률 향상을 도모하기 보다는 조타명령자인 항해사의 조타관련 의도를 예측할 수 있는 파라미터를 적용할 수 있는 후처리 음성인식 프로세스를 개발하였다.

이 프로세스의 성능 검증을 위하여 20개의 조타명령문을 선택한 후 랜덤하게 각각 25회씩 발생하여 실험용 조타명령 구문 데이터베이스를 획득하였다. 그리고 전처리 음성인식 프로세스인 DTW 패턴인식과정과 화자의도예측 인식과정의 인식률 비교실험을 수행하였다.

실험결과 화자의도예측 인식과정을 통한 인식률은 DTW 패턴인식 과정보다 15.4% 향상됨을 알 수 있었다. 특히 화자의도예측 음성인식 과정에서 인식률이 80%이하인 'Port fifteen', 'Starboard ten', 'Ease-to ten'의 경우를 제외한 평균 인식률은 약 94.6%로 나타났다. 따라서 음향학적으로 인식률을 향상시킬 수 있는 표준패턴 학습이론 및 HMM(hidden markof model)과 같은 알고리즘의 적용을 통하여 높은 인식성능이 요구되는 선박

조타명령 음성인식 시스템의 구현이 가능함을 알 수 있었다.

차후 연구 방향으로는 음향학적으로 인식성을 향상시킬 수 있는 알고리즘의 적용과 아울러 과거 입력패턴의 천이확률 해석을 통하여 화자의도예측 파라미터를 보완함으로써 그 성능 개선하고자 한다.

참고문헌

- [1] 박계각, "선박자동화 및 해양안전정보 시스템 현황과 전망", 전자공학회지, 제34권, 제11호, pp.1262-1270, 2007.
- [2] 김종성, 김진권, 문성배, 배병덕, 이형기, 조동욱, 최신헌해실무, 다솜출판사, pp.94-100, 2005.
- [3] 손남선, 김선영, "시뮬레이터 기반 음성을 이용한 항행정보 안내시스템 개발", 한국해양환경 공학회지, 제5권, 제3호, pp.28-31, 2002.
- [4] 이행새, 음성인식, 청문각, pp.6-8, 1997.
- [5] Cole R., Hirschman L., "The Challenge of Spoken Language Systems : Research Directions for the Nineties", IEEE trans. on Speech and Audio Processing, Vol. 3, No. 4, pp.1-18, 1995.
- [6] 최환진, 오영환, "의도종속 문형구조의 학습에 기반한 대화음성의 의도분석", 정보과학회논문지, 제23권 8호, pp.863-864, 1996.
- [7] 문성배, 채양범, 전승환, "조타명령의 음성인식을 위한 최적 특징 파라미터 검출에 관한 연구", 해양환경안전학회지, 제13권 제2호, pp.164-165, 2007.
- [8] 최일홍, 장승관, 차태호, 최웅세, 김창석, "변곡점 및 단구간 에너지평가에 의한 음성의 천이 구간 특징분석", 한국음성과학회지, 제3권, pp.159-160, 1998.
- [9] Rabiner L.R., Juang B.H., "Fundamentals of Speech Recognition", Prentice Hall, pp.112-116, 1999.
- [10] 이기희, 박두석, "음성인식기술과 현황", 한국 OA학회지, 제4권 2호, pp. 2~3, 1996.
- [11] 안태욱, 변용규, 김형협, "MSVQ를 이용한 HMM에 의한 단독어 인식", 전자공학논문지, 제27권, 제9호, pp.158-159, 1990.
- [12] 김희린, "음성인식 기술", 한국멀티미디어학회지, 제7권 2호, pp.16-20, 2003.

저 자 소 개



문성배(文聲培)

1968년생, 1991년 한국해양대학교 항해학과 졸업, 1998년 한국해양대학교 대학원 해사수송과학과 졸업(공학석사), 2002년 한국해양대학교 대학원 해사수송과학과 졸업(공학박사), 현재 한국해양대학교 항해시스템공학부 조교수