

조타명령의 음성인식을 위한 최적 특징파라미터 검출에 관한 연구

문성배* · 채양범** · 전승환***

*,**,*** 한국해양대학교 항해시스템공학부

Optimal Feature Parameters Extraction for Speech Recognition of Ship's Wheel Orders

Serng Bae Moon* · Yang Bum Chae** · Seung Hwan Jun***

*, **,*** Division of Navigation System Engineering, Korea Maritime University, Pusan, 606-791, Korea

요 약 : 이 논문은 선박의 자동조타장치를 음성인식으로 제어할 수 있는 시스템을 개발하기 위한 기초연구로 SMCP(IMO Standard Marine Communication Phrases)에 제시된 조타명령문의 구성 형태를 분석하여 화자의 의도를 예측할 수 있는 특징 파라미터를 추출하였다. 그리고 이 파라미터를 이용하여 1차 패턴인식 과정으로부터 도출된 후보단어 집합으로부터 최종 단어를 결정하는 후처리 인식 프로시저를 설계하였다. 이 프로시저의 유용성을 검증하기 위하여 음성인식용으로 총 525개의 조타명령문을 획득하였고, 표준패턴 기반의 인식과정 인식률과의 비교실험을 수행하였다. 실험결과 의도예측 특징 파라미터를 이용한 인식 프로시저의 인식률이 약 42.3% 향상되어 유효함을 알 수 있었다.

핵심어 : 자동조타장치, 음성인식, 조타명령, 화자의도, 표준패턴

ABSTRACT :The goal of this paper is to develop the speech recognition system which can control the ship's auto pilot. The feature parameters predicting the speaker's intention was extracted from the sample wheel orders written in SMCP(IMO Standard Marine Communication Phrases). And we designed the post-recognition procedure based on the parameters which could make a final decision from the list of candidate words. To evaluate the effectiveness of these parameters and the procedure, the basic experiment was conducted with total 525 wheel orders. From the experimental results, the proposed pattern recognition procedure has enhanced about 42.3% over the pre-recognition procedure.

KEY WORDS : Auto pilot, Speech recognition, Wheel Orders, Speaker intention, Standard pattern

1. 서 론

1990년대 이후 해운선사의 경영합리화, 인건비 절감, 선박자동화 등을 이유로 승무 인원이 감축되어 15~17명만이 승선하는 선박이 일반화되면서 선교 항해당직 업무형태에도 많은 변화가 발생하였다(한국해양대학교 항해실습교재 편찬위원회, 2005). 이러한 변화의 하나로 항해 중 조타수가 선박의 안전순찰, 주간 선박정비 업무의 수행을 위하여 선교를 떠나게 되어 당직 항해사 단독으로 항해당직 업무를 수행하는 이른바 선교당직 1인 체제가 일반화되었다는 것이다(김 등, 2005).

선교 1인 체제에서 당직 항해사는 여러 가지 조산업무와 직접적인 자동조타장치(Auto Pilot)의 제어도 수행하여야 하기 때문에 돌발적인 위급 상황이 발생하게 되면 선박의 위험도는 높아질 수밖에 없는 실정이다. 따라서 대양 또는 연근해와 같이

항행위험도가 높지 않은 수역을 항행할 때, 당직 항해사의 직접적인 조작이 아닌 음성명령을 통하여 자동조타장치를 제어할 수 있는 보조시스템이 있다면 선교당직 1인 체제에 잠재되어 있는 위험도를 낮추는데 크게 기여할 수 있을 뿐만 아니라 갑판부 인력자원을 효율적으로 운용하는데 도움이 될 것이다.

조타명령은 선박의 안전과 직결되는 요소이기 때문에 높은 신뢰성을 가진 인식률은 매우 중요한 개발의 전제조건이라 할 수 있는데, 최근의 음성인식 기술은 지속적인 기술개발에 힘입어 다양한 분야의 실생활에 적용될 수 있는 수준으로 발전하였지만 아직 수많은 응용분야에서 적극적으로 활용되기에는 아직 해결해야 할 기술적 과제가 산적해 있는 실정이다(김, 2003).

특히 선박에서 선교업무의 피로도를 감소시키고, 비상상황시 효과적인 대처를 위하여 음성인식과 음성합성 기법을 적용한 항행정보안내시스템 등에 관한 연구가 이루어졌지만 낮은 음성인식률과 안정성의 문제로 크게 활용되지 못하고 있다(손과 김, 2002).

조타명령의 경우 사용되는 단어의 수가 약 40개 이내로 제한

* 대표저자 : 정희원, mmsbae@hhu.ac.kr, 051)410-4280

** chaeby@hhu.ac.kr, 051)410-4287

*** korjun@hhu.ac.kr, 051)410-4245

되어 있지만 명령문이 한국어가 아닌 영어이기 때문에 화자의 학력수준에 따라 발음에 차이가 발생하고, 또한 발성스타일이 낭독체 또는 자연 음성과 달리 명령어조의 음운 특성을 가지기 때문에 음성인식 환경이 열악한 것이 사실이다(이, 1997).

이와 같은 환경에서 조타명령에 대한 신뢰도를 높이기 위해서는 발성된 단어 또는 음소에 대한 인식률을 향상시킬 수 있는 기법을 적용하여야 할 뿐만 아니라 마이크로폰을 통하여 수신할 때 부가적으로 포함되는 비음성 및 타인의 음성 잡음 등으로 인한 신호대 잡음비의 저하를 개선할 수 있는 아날로그 필터와 같은 하드웨어의 보완이 필수적으로 수반되어야 한다.

한편 화자의 발성 의도를 알 수 있다면 그 의도와 관련된 지식을 음성신호의 음향적 해독이나 구문처리 과정에 적용함으로써 화자가 원하는 것을 수행하게 하는 음성이해(Speech Understanding)의 처리기법이 있는데, 최근 화자의 의도에 기반하여 음성인식의 오류를 줄이려는 연구가 활발히 진행되고 있다(Cole and Hirschman, 1995 ; 최와 오, 1996).

이 논문에서는 음향학적 인식기법의 적용 또는 하드웨어의 최적화를 통하여 인식성능을 향상시키기 보다는 조타명령 구문의 분석을 통하여 화자의 발성의도를 어느 정도 예측할 수 있는 특징 파라미터를 추출하여 인식률을 향상시킬 수 있는 후처리 인식과정을 개발하는 것을 목적으로 기초적 연구를 수행하였다.

그리고 이 연구의 최종목표는 항해사의 조타명령을 실시간으로 인식하고 자동조타기를 제어할 수 있는 임베디드 시스템(Embedded System) 기반의 음성인식 자동조타 시스템을 개발하고자 하는 것이다.

2. 조타명령 음성인식시스템

Fig.1은 이 연구에서 기초실험을 위하여 구현한 조타명령 음성인식 시스템의 전체적인 흐름을 도시한 것으로서, 크게 음성신호 입력부(Microphone, Amp., Analog Filter, AD Converter), 음성신호 검출부(Speech Detection), 특징 추출부(Feature Extraction, Target Pattern), 벡터양자화(Vector Quantization, Reference Pattern) 및 인식부(Pattern Recognition, Speaker Intention Inference Procedure, Decision)로 나눌 수 있다.

2.1 음성신호 입력

음성신호 입력부는 마이크로폰, 증폭기, 아날로그 필터 및 AD변환기로 구성하였다. 마이크로폰은 범용의 PC용 헤드셋에 장착된 콘덴서형의 무지향성 특성을 가진 것으로 선정하였고, LM358 OP Amp. 소자를 이용하여 증폭기와 아날로그필터를 제작하였다.

증폭기의 증폭률은 약 38dB이고, 아날로그 저역통과필터의 차단주파수는 3.5kHz로 설계하였다. Table 1은 AD변환기의 주요 사양을 나타낸 것이다.

2.2 음성신호 검출

발성모드에 따른 음성인식 시스템은 크게 단어 사이의 경계를 나타내기 위하여 휴지기가 요구되는 고립단어인식(Isolated Word Recognition)과 경계를 나타내기 위한 명시적 조건이 필

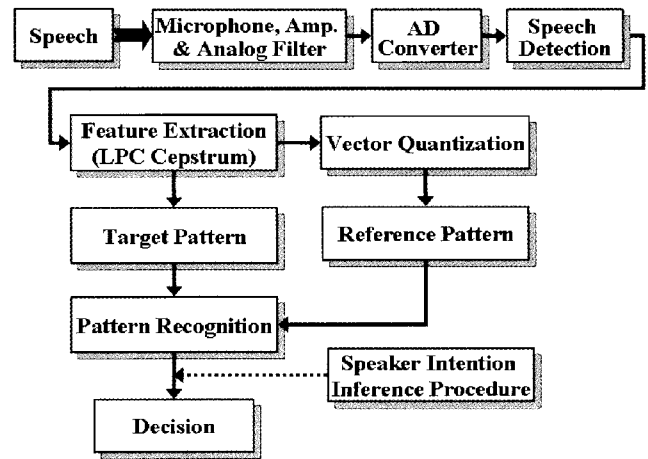


Fig. 1. Block Diagram of Recognition System.

Table 1. Specifications of AD Converter

Maker	TiePie Engineering, Netherlands
Model	Handyscope 3
Interface	USB 2.0 & USB 1.1
Channels	AD : 2, DA :1
Input Sensitivity	200mV-80V full scale
Input Impedance	1Mohm/30pF
Resolution	12bits
Sampling Frequency	8kHz
Sampling Period /Sample Number	3.75sec / 30,000samples

요 없는 연속음성인식(Continuous Speech Recognition)으로 나눌 수 있다(이와 박, 1997 ; 오, 1998). 이 논문에서는 조타명령에 사용되는 어휘의 수가 많지 않고 발음에 연속성이 비교적 적은 것을 고려하여 고립단어인식 기술을 적용하였다.

AD변환기를 통하여 획득한 전체 신호데이터에 대하여 음성 부분만을 검출하는 기법으로 주어진 단구간 내에서 기준값인 0을 통과하는 횟수를 측정하는 영교차율(Zero-crossing Rate)과 단구간 내의 각 샘플을 자승하여 합을 구한 단구간 에너지 합수법이 있다(이, 1997). 이 연구에서는 유성음과 무성음의 구별에 이용될 뿐만 아니라 고음질의 신호에서 목음과 무성음의 구별에 유용하게 이용될 수 있는 단구간 에너지 합수법을 적용하였다. 식(1)은 N개의 샘플을 가진 단구간에 대한 에너지를 나타낸 것이다(최 등, 1998).

$$E(m) = \sum_{n=m-N+1}^m [s(n)u(m-n)]^2 \quad (1)$$

단, $s(n)$ 은 샘플값, $u(n)$ 은 단구간 해밍창함수,
 m 은 각 구간의 N 번째 샘플

2.3 특징추출

음성인식을 위하여 사용되는 특징벡터 추출 기법으로는 LPC(Linear Predictive Coding) 캡스트럼, PLP(Perceptual Linear Prediction) 캡스트럼, MFCC(Mel Frequency Cepstral Coefficient), Filter Bank Energy 등이 있지만 어느 것이 좋다고 단정할 수 없는 것이 현실이다. 이 연구에서는 음성발생기관의 모델에 근거를 두고 화자중속 음성인식 기법에 많이 사용되는 LPC 캡스트럼을 선택하였다.

여기서 LPC 캡스트럼은 LPC계수 분석을 통하여 구해지는 스펙트럼 포락선의 캡스트럼 계수를 나타내는 것으로서, Levinson-Durbin 알고리즘을 이용하여 LPC 계수를 연산한 후 캡스트럼 계수를 연산하도록 프로그래밍 하였다. 특징 벡터열인 캡스트럼 계수 c_m 는 식(2.a)와 식(2.b)로 나타낼 수 있다 (Rabin and Juang,1999).

$$c_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m} \right) c_k a_{m-k}, \quad 1 \leq m \leq p \quad (2.a)$$

$$c_m = \sum_{k=m-p}^{m-1} \left(\frac{k}{m} \right) c_k a_{m-k}, \quad p < m \quad (2.b)$$

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

2.4 벡터양자화 및 패턴인식

음성인식은 표준패턴과 입력패턴과의 유사한 정도를 측정하는 것이기 때문에 인식대상 어휘별로 기준이 되는 표준패턴이 준비되어야 한다. 이 준비과정을 벡터학습 또는 훈련과정이라고 하는데, 각 어휘별로 추출된 다수의 특징 벡터열을 벡터 양자화하여 표준패턴에 해당하는 코드북을 작성하는 것이다(이와 박, 1996).

벡터 양자화를 이용하여 코드북을 작성하는 방법으로 두 가지가 있다.

첫 번째는 한 단어당 표준패턴으로 하나의 코드북만을 취하는 단일 섹션 코드북(Single section codebook) 작성법이고, 두 번째는 한 단어당 2개 이상의 코드북으로 표준패턴을 작성하는 멀티 섹션 코드북(Multi-section codebook) 작성법이다. 일반적으로 단일 섹션 코드북 작성법은 멀티 섹션 코드북 작성법에 비해 시스템을 작게할 수 있다는 장점이 있지만 인식률이 낮기 때문에 소규모 인식시스템에 제한적으로 적용된다(안 등, 1990).

이 연구에서는 인식률을 높일 수 있는 음성인식 기법을 적용하는 것이 아니고, 특수 목적으로 사용되는 조타명령문의 특징 파라메타가 인식 성능을 어느 정도 개선할 수 있는지를 평가하기 위한 것이기 때문에 단일 섹션 코드북 작성법을 적용하여 음성인식 시스템을 구축하였다.

한편 벡터 양자화 방법에는 K-means, LBG, LVQ 등이 있다. 이 연구에서는 군집화(Clustering) 문제에 있어서 미리 정해

진 K개의 클러스터에 따라 벡터열을 분류하기 위한 것으로 가장 간단한 자율학습(Unsupervised Learning) 알고리즘의 하나인 K-means를 통하여 벡터양자화를 수행하였다(김, 2003).

인식과정에서는 인식하고자 하는 입력음성의 특징벡터인 시험패턴과 각 단어의 표준패턴의 유클리드 거리(Euclidean distance)를 연산하여 그 유사도를 측정하였다. 유클리드 거리는 벡터의 각 요소가 서로 상관관계가 없으면서 분산이 같은 경우에 사용할 수 있는 것으로 K-means 군집화 과정에도 적용하였다(오,1998). 유클리드 거리 d 는 식(3)과 같이 나타낼 수 있다(Baek et al., 2004).

$$d = \sum_{i=1}^n (x_i - y_i)^2 \quad (3)$$

단, x_i, y_i 는 두 벡터열, n 은 벡터의 개수

이 연구에서는 유클리드 거리를 측정하여 시험패턴과 유사도가 높은 후보 단어를 결정하는 패턴인식을 1차 인식과정으로 하였고, 그 후 조타명령 발생자의 의도를 예측할 수 있는 특징 파라미터를 이용하여 최종 인식이 이루어지도록 음성인식 시스템을 설계하였다.

3. 조타명령 구문분석

3.1 조타명령 단어와 구문

IMO의 SMCP(IMO Standard Marine Communication Phrases)에 제시된 조타명령 구문의 종류는 약 26개이고, 이 구문에서 사용되는 단어의 수는 약 40개이다. 이 중에서 자동조타장치 제어에 있어서 조타수와 같은 인간의 지적 판단이 요구되는 동작과 관련된 명령, 즉 "Steady", "Report if she dose not answer wheel", "Keep buoy on port side" 등을 제외하면 19개의 구문형태와 19개의 단어로 축약할 수 있다.

조타명령은 크게 타각(Rudder Angle)을 지정하는 타각지정 명령과 선수방위의 변경을 요청하는 침로지정 명령으로 분류할 수 있다. Table 2는 이 음성인식 시스템에서 인식의 대상으로 선정된 단어의 범주를 나타낸 것이다. Table 2에서 'Ease-to'는 'Ease'와 'To'로 각각 분리되어 두개의 고립단어로 인식되어야 하지만 실제로는 휴지시간이 거의 없이 연속적으로 발생될 뿐만 아니라 'To'가 많이 약화되기 때문에 이 연구에서는 하나의 단어로 취급하여 처리하였다.

Table 3은 벡터 양자화를 통하여 작성된 표준패턴들 사이의 유클리드 거리를 측정한 결과를 나타낸 것으로서 값이 작을수록 두 패턴 간에 유사도가 높다는 것을 의미한다. 그러나 Table 3에 나타난 각 단어 사이의 유클리드 거리는 화자의 발생특성 및 군집화 기법 등에 따라 다르게 측정될 수 있다. 이 표에서와 같이 'One'은 'Five', 'Two', 'Four' 및 'Hard'는 'Port', 'Three'는 'Twenty', 'Five'와 'Starboard'는 'Nine',

Table 2. Words used for Wheel Orders

분류	단어원형
타각지정명령	Midships, Port, Starboard, Hard, Ease-to, Ten, Fifteen, Twenty
공통	Five
침로지정명령	One, Two, Three, Four, Six, Seven, Eight, Nine, Zero, Steer

여된 파라미터의 값을 나타낸 것이다. 예를 들면 'Midships'은 1단어 명령문에서만 발생될 수 있기 때문에 1단어 명령문을 제외한 나머지 명령문에서는 파라미터의 값이 0으로 설정된다. 그리고 'Five', 'Ten', 'Fifteen' 및 'Twenty'를 제외한 숫자는 4단어 명령문에서만 발생될 수 있기 때문에 4단어 명령문에서만 파라미터의 값을 1로 설정하였다. Table 6은 조타명령 구문에서 단어의 위치에 따라 인식 대상 단어별로 부여된 파라미터

Table 3. Normalized Euclidean Distance Matrix between the Order's Patterns

word	One	Two	Three	Four	Five	Six	Seven	Eight	Nine	Ten	Zero	Fifteen	Twenty	Starboard	Port	Hard	Midships	Steer	Ease-to
One	0.0	0.071	0.291	0.132	0.059	0.376	0.195	0.556	0.086	0.285	0.161	0.374	0.275	0.073	0.103	0.137	0.352	0.680	0.420
Two	0.071	0.0	0.199	0.076	0.194	0.262	0.211	0.429	0.196	0.258	0.144	0.256	0.206	0.148	0.068	0.176	0.255	0.522	0.306
Three	0.291	0.199	0.0	0.274	0.377	0.117	0.130	0.168	0.254	0.098	0.100	0.099	0.081	0.200	0.295	0.344	0.137	0.289	0.151
Four	0.132	0.076	0.274	0.0	0.240	0.273	0.212	0.425	0.267	0.277	0.218	0.291	0.258	0.163	0.051	0.216	0.293	0.543	0.390
Five	0.059	0.194	0.377	0.240	0.0	0.497	0.199	0.720	0.035	0.282	0.196	0.514	0.364	0.060	0.175	0.132	0.475	0.903	0.550
Six	0.376	0.262	0.117	0.273	0.497	0.0	0.138	0.082	0.401	0.113	0.149	0.052	0.061	0.337	0.389	0.533	0.026	0.167	0.095
Seven	0.195	0.211	0.130	0.212	0.199	0.138	0.0	0.305	0.132	0.050	0.036	0.198	0.098	0.116	0.255	0.289	0.146	0.462	0.251
Eight	0.556	0.429	0.168	0.425	0.720	0.082	0.305	0.0	0.594	0.246	0.339	0.077	0.167	0.490	0.565	0.737	0.143	0.058	0.136
Nine	0.086	0.196	0.254	0.267	0.035	0.401	0.132	0.594	0.0	0.192	0.129	0.420	0.276	0.044	0.207	0.155	0.378	0.764	0.445
Ten	0.285	0.258	0.098	0.277	0.282	0.113	0.050	0.246	0.192	0.0	0.080	0.157	0.093	0.189	0.301	0.313	0.135	0.416	0.211
Zero	0.161	0.144	0.100	0.218	0.196	0.149	0.036	0.339	0.129	0.080	0.0	0.192	0.092	0.119	0.228	0.252	0.138	0.480	0.232
Fifteen	0.374	0.256	0.099	0.291	0.514	0.052	0.198	0.077	0.420	0.157	0.192	0.0	0.039	0.331	0.401	0.556	0.050	0.116	0.025
Twenty	0.275	0.206	0.081	0.258	0.364	0.061	0.098	0.167	0.276	0.093	0.092	0.039	0.0	0.220	0.341	0.441	0.031	0.235	0.048
Starboard	0.073	0.148	0.200	0.163	0.060	0.337	0.116	0.490	0.044	0.189	0.119	0.331	0.220	0.0	0.123	0.115	0.325	0.663	0.384
Port	0.103	0.068	0.295	0.051	0.175	0.389	0.255	0.565	0.207	0.301	0.228	0.401	0.341	0.123	0.0	0.073	0.412	0.742	0.506
Hard	0.137	0.176	0.344	0.216	0.132	0.533	0.289	0.737	0.155	0.313	0.252	0.556	0.441	0.115	0.073	0.0	0.558	1.000	0.663
Midships	0.352	0.255	0.137	0.293	0.475	0.026	0.146	0.143	0.378	0.135	0.138	0.050	0.031	0.325	0.412	0.558	0.0	0.187	0.060
Steer	0.680	0.522	0.289	0.543	0.903	0.167	0.462	0.058	0.764	0.416	0.480	0.116	0.235	0.663	0.742	1.000	0.187	0.0	0.135
Ease-to	0.420	0.306	0.151	0.390	0.550	0.095	0.251	0.136	0.445	0.211	0.232	0.025	0.048	0.384	0.506	0.663	0.060	0.135	0.0

'Six'와 'Twenty'는 'Midships', 'Seven'은 'Zero', 'Eight'은 'Steer', 'Nine'은 'Five', 'Ten'과 'Zero'는 'Seven', 'Fifteen'은 'East-to', 'Port'는 'Four', 'Steer'는 'Eight', 'Ease-to'는 'Fifteen'과 유사도가 높은 것을 알 수 있었다.

3.2 의도예측 파라미터

조타명령의 경우는 SMCP에 따라 특정 상황에서 한정되어 이용되기 때문에 음향학적 해석에 관계없이 조타 명령문을 구성하는 단어 개수와 단어의 위치(순서)와 같은 파라미터에 따라 화자의 의도예측이 어느 정도 가능하다고 할 수 있다.

Table 4는 SMCP를 바탕으로 이 연구에서 인식의 대상으로 선정한 조타명령문의 기본 형태와 종류를 나타낸 것으로서, 크게 타각지정명령 구문과 침로지정명령 구문으로 나눌 수 있다. 또한 구문을 형성하는 단어의 수에 따라 1단어 명령문, 2단어 명령문, 3단어 명령문 및 4단어 명령문으로 분류할 수 있다.

Table 5는 이 명령문의 형태에 따라 인식 대상 단어별로 부

값을 나타낸 것이다. 특히 침로지정 명령문에서 침로각은 360° 방식으로 지정되기 때문에 0에서 9까지의 숫자 중에서 'Zero', 'One', 'Two', 'Three'를 제외한 'Four', 'Six', 'Seven', 'Eight', 'Nine'은 두 번째 위치에서 발생될 수 없는 단어로 간주하여 파라미터 값을 결정하였다. 그러나 'Five'는 타각지정 명령문에서 두 번째 단어로 이용될 수 있기 때문에 발생가능 파라미터로 분류하였다.

4. 실험 및 평가

4.1 실험방법

이 연구에서 구현한 음성인식 시스템은 화자종속방식을 근간으로 하고 있기 때문에 음성인식 대상 단어에 대한 표준패턴의 작성에 필요한 훈련용 음성 DB와 조타명령문의 발생자로 39세이 남성 1인을 선정하였다. 그리고 실험은 보통의 조용한

Table 4. Standard Structure of Wheel Orders

분 류	구 문	구문 구성 단어수
타각지정명령	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
	Port twenty five	3
Starboard twenty five	3	
침로지정명령	Steer zero seven zero	4
	Steer one two five	4
	Steer two four nine	4
	Steer three six eight	4

Table 5. Parameters Depending on Word Number

단어	1단어 명령문	2단어 명령문	3단어 명령문	4단어 명령문
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

실내에서 수행하였다.

Table 2의 인식 대상 단어에 대한 표준패턴은 각 단어를 20회씩 발성하도록 한 후 12차 LPC 웨스트럼 계수를 연산하여 특징 벡터 열을 추출한 후 벡터 양자화를 통하여 작성하였다. 그리고 Table 4의 실험용 조타 명령문을 동일화자가 랜덤하게 각각 25회씩 발성되도록 하여 총 525개의 조타명령 음성데이터를 수집하였다.

한편 이 음성인식 시스템은 벡터양자화로 구축한 표준패턴과 시험패턴의 유클리드 거리를 측정하고, 각 단어별로 4개의 1차 후보단어 집합을 결정하도록 하였다. 여기서 후보단어 집합의 원소 개수는 Table 2에서 유클리드 거리가 0.01이하로 측정된 경우의 수를 평균한 값이 4.1579이기 때문에 후보단어의 수는 4개로 결정하였다. 그리고 후보단어 집합에 대하여 Table 5와 Table 6의 조타명령 구문 특징 파라미터 값을 근거로 의도

Table 6. Parameters Depending on Word Position

단어	첫 번째 단어	두 번째 단어	세 번째 단어	네 번째 단어
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

예측 추론과정이 수행되어 최종적으로 인식단어가 결정되도록 시스템을 프로그래밍하였다.

4.2 의도에측 패턴인식 실험결과

화자의 의도를 예측할 수 있는 특징 파라미터를 근거로 음성인식을 수행하는 패턴인식 과정의 유용성을 검증하기 위하여 표준패턴과 입력패턴의 유사도만으로 단어를 결정하는 1차 패턴인식 과정과의 비교실험을 수행하였고 그 결과는 Table 7과 같다.

먼저 1차 패턴인식 과정에 의한 평균 인식률은 Table 6에서와 같이 약 32.9%로 매우 낮게 나타나 실제 선박에서 조타기를 제어할 수 있는 음성인식 시스템으로 적용하기에는 거의 불가능함을 알 수 있다. 이것은 K-means 알고리즘을 바탕으로 단일 섹션 코드북 형태의 표준패턴 구축과 패턴인식이 수행되어 화자의 물리적 또는 감정적인 상태에 따라 발성속도(Speaking Rate)와 음질의 변화로 야기되는 화자내부 변이(With-speaker Variability)와 음성변이(Phonetic Variability) 등의 문제가 해결되지 못한 것이 원인이라 할 수 있다. 그러나 이러한 문제는 인식률 성능을 높일 수 있는 특징 벡터 추출법과 동적시간정합(DTW, Dynamic Time Warping) 또는 은닉마르코프모델(HMM, Hidden Markov Model) 등과 같은 인식 알고리즘을 적

용하면 해결할 수 있는 것이다.

한편 의도예측 추론용 특징 파라미터를 이용한 경우는 평균 인식률이 약 75.2%로 나타나 1차 패턴인식 과정보다 약 42.3% 향상됨을 알 수 있다. 특히 'Ease-to five' 명령문의 경우 1차 패턴인식 과정을 통과할 경우 25회의 발성 중에서 단 2회만이

Table 7. Recognition Results

명령문	1차 패턴인식	의도예측 특징 파라미터 패턴인식
Midships	100%	100%
Port five	44%	100%
Port ten	28%	80%
Port fifteen	20%	96%
Port twenty	24%	48%
Hard port	8%	84%
Starboard five	32%	100%
Starboard ten	16%	36%
Starboard fifteen	28%	68%
Starboard twenty	44%	80%
Hard Starboard	40%	72%
Ease-to five	8%	100%
Ease-to ten	76%	76%
Ease-to fifteen	68%	96%
Ease-to twenty	60%	92%
Port twenty five	4%	60%
Starboard twenty five	8%	80%
Steer zero seven zero	20%	60%
Steer one two five	44%	72%
Steer two four nine	12%	44%
Steer three six eight	8%	36%
평균	32.9%	75.2%

정상적으로 인식되었지만 특징 파라미터를 이용한 경우는 25회 모두 정상적으로 인식되는 변화를 보였다. 그리고 침로지정 명령문인 'Steer zero seven zero', 'Steer one two five', 'Steer two four nine' 및 'Steer three six eight'의 인식률은 평균 53%로 나타나 그 나머지 명령문에 대한 인식률 약 80.5%에 비해 인식률이 낮음을 알 수 있었다. 이것은 침로지정 명령문의 주요 구성 단어인 숫자가 다른 단어에 비해 화자의 의도를 예측할 수 있는 특징이 적기 때문이다.

실험결과 화자의 의도를 예측할 수 있는 특징 파라미터를 이용한 음성인식의 경우 인식률이 크게 향상되었지만 실용화의 측면에서 볼 때는 아직 그 성능이 좋지 않음을 알 수 있었다. 그러나 벡터양자화 코드북을 작성하기 위한 훈련 DB의 추출방법의 변경과 화자내부 변이 및 음성변이의 영향을 최소화할 수 있는 알고리즘의 적용을 통하여 조타명령 단어의 인식률을 향상시킬 수 있다면, 의도예측 특징 파라미터를 이용한 음성인식 엔진의 성능은 크게 향상되어 실용화를 도모할 수 있을 것으로 기대된다.

5. 결론

이 연구에서는 조타명령 발성시 내포된 화자의 의도를 예측하여 발성음의 음향학적 패턴인식 과정을 보정할 수 있는 특징 파라미터를 추출하였다. 그리고 이 파라미터를 이용한 패턴인식 과정의 성능을 검증하기 위하여 525개의 음성 DB를 획득하여 기초 실험을 수행하였다. 실험결과 의도예측 추론용 특징 파라미터를 이용한 인식과정의 인식률은 약 75.2%이었고, 1차 패턴인식 과정의 경우는 약 32.9%로 측정되어 약 42.3%정도 인식률이 향상됨을 확인할 수 있었다. 따라서 음향학적으로 인식률을 향상시킬 수 있는 DTW와 HMM와 같은 여러 가지 알고리즘이 적용된다면 매우 높은 인식률이 요구되는 선박 조타명령 음성인식 시스템의 구현이 가능함을 알 수 있었다.

앞으로의 연구 방향으로는 DTW 또는 HMM의 적용을 통한 음성변이 및 화자내부 변이 문제를 해결하고, 과거 입력패턴의 천이확률 해석을 통하여 의도예측 특징 파라미터를 보완하고자 한다.

후 기

이 연구는 한국해양대학교 신진교수 연구지원사업의 지원을 받은 과제입니다.

참 고 문 헌

- [1] 김종성, 김진권, 문성배, 배병덕, 이형기, 조동욱(2005), 최신항해실무, 다솜출판사, pp. 94~100.
- [2] 김희린(2003), 음성인식 기술, 한국멀티미디어학회지, 제7권 2호, pp. 16~20.
- [3] 손남선, 김선영(2002), 시뮬레이터 기반 음성을 이용한 항행정보 안내시스템 개발, 한국해양환경공학회지, 제5권, 제3호, pp. 28~31.
- [4] 안태욱, 변용규, 김형협(1990), MSVG를 이용한 HMM에 의한 단독어 인식, 전자공학논문지, 제27권, 제9호, pp. 158~159.
- [5] 오영환(1998), 음성언어정보처리, 흥릉과학출판사, pp. 91~93.
- [6] 이기희, 박두석(1996), 음성인식기술과 현황, 한국OA학회지, 제4권, 제2호, pp. 2~3.
- [7] 이형새(1997), 음성인식, 청문각, pp. 6~8.
- [8] 최일홍, 장승관, 차태호, 최웅세, 김창석(1998), 변곡점 및 단구간 에너지평가에 의한 음성의 천이구간 특징분석, 한국음성과학회지, 제3권, pp. 159~160.
- [9] 최환진, 오영환(1996), 의조종속 문형구조의 학습에 기반한 대화음성의 의도분석, 정보과학회논문지, 제23권, 제8호, pp. 863~864.
- [10] Baek S.J., Kim J.Y., Kang S.K.(2004), An Approximate Euclidean Distance Calculation for Fast VQ Encoding, The Korean Association of Speech Science, Vol. 11, No. 2, pp. 211~213.
- [11] Cole R., Hirschman L.(1995), 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.

[12] Rabiner L.R., Juang B.H.(1999), Fundamentals of
Speech Recognition, Prentice Hall, pp. 112~116.

원고접수일 : 2007 년 5 월 14 일

원고채택일 : 2007 년 6 월 26 일