

Mahalanobis Taguchi System

홍 정 의[†]

충주대학교 공과대학 산업경영공학과

Diagnosis of Parkinson's Disease by Voice Disorder Using Mahalanobis Taguchi System

Jungeui Hong[†]

Department of Industrial and Management Engineering, Chungju National University

Human voice reacts very sensitively to human's minute physical condition. For instance, human voice disorders affect patients profoundly especially in the case of Parkinson's disease. Acoustic tools such as MDVP, can function as an equipment that measures various voice in different objects. Many different approaches have been applied for analyzing the voice disorders for diagnosis of Parkinson's disease. According to the voice data of suspected Parkinson's patients from UCI Machine Learning Repository, it is reported to have 23 people with Parkinson's disease and 8 healthy people. Applying Mahalanobis Taguchi System (MTS) for diagnosis of Parkinson's disease, the correct diagnosis performance is compared to previous research results.

Keywords : MTS (Mahalanobis Taguchi System), Parkinson's disease, Diagnosis

1. 서 론

인간의 질병을 보다 효과적이고 경제적으로 진단하기 위해서는 질병의 진단에 꼭 필요한 검사항목을 설정하고 이를 정확하게 분석하는 일이 무엇보다 중요하다. 이러한 과정은 다변량 시스템의 해석절차와 같은 절차를 통해 해석할 수 있다. 자동차의 브레이크 설계, 유방암의 진단 그리고 음성분석 통한 파킨슨병의 진단 등은 다수의 변수를 가지고 변수간의 상호작용을 배제할 수 없는 비선형 다변량 시스템의 해석을 통해 해석할 수 있다. 대부분의 다 변량 시스템은 변수들의 변화에 따른 측정값의 변화를 얼마나 정확히 예측하는가를 그 분석의 기본으로 삼고 있다. 그러나 일반적으로 이러한

해석을 위해서는 많은 양의 데이터가 필요로 한다.

사람의 말소리는 미세한 몸의 상태변화에 민감하게 반응하면 연령 또는 건강상태에 따라 변하기도 한다. 파킨슨병은 신경전달물질인 도파민의 결핍으로 신체의 운동기능이 퇴행하는 대표적인 신경계 퇴행성 질환이다. 사지 근육경직, 휴지기 진전, 운동 완만 또는 자세 불안정의 특징적인 임상증상을 보인다. 또한 병이 진행될수록 글씨의 크기가 작아지고, 얼굴표현이 감소하며, 불분명한 어투 등의 증상이 나타나기도 한다. 특히 파킨슨병에서 보이는 구어 장애로는 입술과 턱 위치의 강직으로 인해 조음범위의 축소로 인하여 음성의 정확도가 감소하는 조음장애, 호흡문제와 진전으로 인한 단조로운 음도, 거친 음성, 성대 떨림 등의 음성장애를 들

수 있다[1, 2, 3].

그동안 음성분석을 통한 파킨슨병의 진단해석을 위해 수많은 연구가 이루어져 왔으며 보다 정확한 해석을 위해 그 방법들은 nonlinear dynamical system theory, neural network, radial basis function 등 복잡한 해석방법[4, 5, 6, 7]들이 사용되어지고 있다.

본 연구는 기존의 음성연구데이터를 Mahalanobis Taguchi System을 이용하여 진단한 결과를 기존의 진단방법과 비교하고 나아가 최적화 과정을 통해 불필요한 진단 변수를 제거하고 파킨슨병의 진단에 큰 영향을 미치는 변수만을 이용하여 진단에 활용함으로써 보다 효율적인 진단방법을 제시하고자 한다.

2. Mahalanobis Taguchi System

마할라노비스 거리(Mahalanobis Distance : MD)는 인도의 수학자 Mahalanobis에 의해 한 집단에서 이질의 집단을 구분하는 방법으로 1930년대에 소개되어 졌다. 강건 설계 방법을 고안해낸 다구찌는 어떤 집단의 평균값을 기초로 한 마할라노비스 공간(Mahalanobis Space : MS)를 설정하고 이를 기초로 한 새로운 관측 값이 이러한 공간으로부터 얼마나 벗어나 있는가를 측정하는 마할라노비스 다구찌 시스템(Mahalanobis Taguchi System : MTS) 방법을 고안해 냈다. MTS에서는 다차원의 단위 공간으로서 MS를 정의 하고 임의의 대상이 그 공간으로부터 얼마만큼 떨어져 있는가를 나타내는 방법으로 MD를 이용한다. 또한 설정된 측정공간으로부터의 진단의 정확도를 평가하기 위하여 다구찌 방법이 사용되어 진다[8, 9]. 이러한 MTS 방법의 장점은 다변량 함수해석에 매우 중요한 변수들 간의 상관관계를 고려한다는 것이다.

MTS 기법은 다차원의 공간으로부터 MS를 정의 하고 임의의 측정대상이 그 공간으로부터 얼마만큼 떨어져 있는가를 데이터 해석학적인 방법으로 분석해 내는 해석 방법이다. MTS 방법을 이용한 정확한 예측을 위해서는 다차원의 공간을 대표하는 단위공간을 설정하는 것이다. 다차원 공간에서의 관측된 패턴은 변수들 간의 상호 상관관계에 영향을 받으며 종종 이러한 상관관계를 무시하고 독립된 변수로만 생각하여 잘못된 해석을 내리는 경우가 많이 발생하고 있다.

MTS에서 MS는 정상 또는 건강한 그룹의 표준화된 변수들을 이용하여 구할 수 있으며 이를 이용하여 건강한 그룹과 건강하지 못한 그룹을 구분하는 지표로 삼을 수 있다. MS가 구해지면 변수들 중 측정치에 영향을 미치는 정도를 판단하기 위하여 S/N비(Signal to Noise Ratio)와 직교배열표(Orthogonal Array : OA)를 이용한다.

MTS의 일반적인 적용절차는

첫째 표준이 되는 집단으로부터 판단에 적용될 변수들을 선정한다. 여기서 선정된 변수들에 의해 구성된 MS를 구성한다. 이를 위하여 정상 또는 건강한 그룹의 데이터가 사용되며 계산된 마할라노비스 거리의 평균 값은 1에 근접한다.

$$z_i = \frac{X_i - m}{\sigma} \quad (1)$$

여기서 m 은 변수의 평균 값이고,

σ 는 표준편차 그리고 X_i 는 임의 측정값이다.

다차원 공간에서의 마할라노비스 거리는 변수들 간의 상관관계를 계산함으로써 구할 수 있다. 이러한 MD의 통계학적인 의미는 임의의 측정값이 선택 집단의 중간 값으로 부터 얼마나 근접해 있는가를 의미한다. 아래의 공식은 마할라노비스 거리를 계산하는 공식이다.

$$MD_j = D_j^2 = \frac{1}{k} Z_{ij}^T C^{-1} Z_{ij} \quad (2)$$

여기서 C^{-1} 은 변수들 간의 상관계수를 포함한 상관행렬의 역행렬이고 T는 표준 벡터의 transpose 벡터이다. 선택된 MS 공간에서 구해진 MD 값의 평균값은 대략 1에 근접한다. 따라서 이러한 MS 공간을 단위공간이라고 부른다.

두 번째는 이렇게 구해진 MS 공간의 유효성을 판단한다. 이를 위해서 MS 공간 밖의 측정값 즉 비정상 또는 건강하지 않은 데이터를 이용한다. 이러한 집단의 MD를 구하기 위해서 정상 또는 건강한 집단으로 단위 MS를 구성하는 평균, 표준편차 그리고 상관행렬을 이용한다. 단위 공간이 유효하다면 비정상 또는 건강하지 않은 집단의 MD 값은 정상 또는 건강한 집단의 MD 값보다 훨씬 커서 구별이 뚜렷할 것이다.

세 번째 과정은 변수 중에서 측정값에 영향을 미치지 않거나 적게 미치는 변수를 찾아내서 제거 하여 시스템의 해석을 쉽게 하는 일이다. 이러한 목적을 위하여 OA와 SN비가 유용하게 활용되어 질수 있다. 직교배열표의 열은 실험변수들을 배열 하였고 행은 실험의 조합을 나타낸다. 즉 직교 배열표의 Level 1은 변수를 사용하는 경우를 의미 하고 Level 2는 변수를 사용하지 않는 경우를 의미 한다. 따라서 직교배열표의 조건에 따라 변수들은 사용되어지거나 무시될 수 있으며 이를 바탕으로 SN비를 계산 할 수 있다. 대당 특성의 SN비를 구하는 식은

$$SNration = \eta = -10 \log \left(\frac{1}{t} \sum_{i=1}^t \frac{1}{D_i^2} \right) \quad (3)$$

여기서 η 는 S/N ratio이고 t 는 비정상 그룹의 개수이며, D_i^2 는 i 번째의 마할라노비스 거리이다.

3. MTS를 이용한 환자의 음성분석을 통한 파킨슨 병의 진단

파킨슨병 환자의 언어장애 특징은 전반적으로 음성이 힘이 없어지고, 단조로워 지며, 연속적으로 발음할 때 음성의 강도변화가 점차적으로 감소하며, 발음이 정확하지 못한 특징이 있다. 이는 근육이 경직되어 발생하는 것으로 알려져 있다. 본 연구는 파킨슨 증상을 나타내는 환자의 음성 데이터를 분석하여 각각의 진단 항목과 파킨슨 병과의 상관관계를 관찰하고 나아가 효과적인 음성분석을 통한 파킨슨병의 진단방안을 제시하고자 한다. Mahalanobis Taguchi System은 다수의 변수를 포함하는 이러한 다변량 시스템의 해석에 아주 효과적으로 활용되어져 왔고[10, 11, 12, 13] 최적화 해석 과정을 통해 보다 경제적이고 효과적으로 새로운 진단의 방법을 제시 할 수 있을 것으로 판단된다.

3.1 변수 데이터

본 연구를 위한 음성데이터는 UCI machine learning

<표 1> 데이터의 변수 항목

변 수	항 목
A1	MDVP Fo Hz(Average Vocal Fundamental Frequency)
A2	MDVP Fhi Hz(Maximum Vocal Fundamental Frequency)
A3	MDVP Flo Hz(Minimum Vocal Fundamental Frequency)
A4	MDVP Jitter %
A5	MDVP Jitter Abs
A6	MDVP RAP
A7	MDVP PPQ
A8	Jitter DDP
A9	MDVP Shimmer
A10	MDVP Shimmer(dB)
A11	Shimmer APQ3
A12	Shimmer APQ5
A13	MDVP APQ
A14	Shimmer DDA
A15	NHR(Noise-to-Harmonic Ratio)
A16	HNR(Harmonics-to-noise Ratio)

MDVP-Muti-Dimensional Voice Program.

depository의 음성데이터를 이용하였으며, 사용된 음성 데이터는 23명의 파킨슨 환자의 음성과 8명의 정상인의 음성을 <표 1> 같은 항목을 변수로 하여 6번 또는 7번 반복 측정 48개의 정상인 데이터 그리고 147개의 파킨슨병 환자의 데이터를 포함한 총 195개의 데이터를 가지고 분석을 시행 하였다.

데이터는 먼저 정상인의 데이터와 파킨슨환자의 데이터를 분리하여 각각의 MD를 계산하여 이상치(Outlier)로 여겨지는 데이터를 제거하였다. 정상인그룹에서는 MD값이 1.5이상인 8개의 데이터를 이상치로 판단하였으며, 파킨슨 환자의 데이터에서는 MD값이 3이상인 6개의 데이터를 이상치로 판단하였다. 다음은 진단을 위한 MS설정을 위해 정상인그룹에서 38개의 데이터 그리고 파킨슨 환자의 그룹에서 13개의 데이터를 임의로 선택하여 정상그룹(Normal)과 비정상 그룹(Abnormal)으로 분리 하였다.

선택된 정상 그룹과 비정상그룹의 음성데이터는 각각 <표 2>와 <표 3>과 같다.

3.2 MTS기법의 전개

<표 2> 정상그룹을 식 (1)을 이용하여 표준화과정을 통해 상관행렬을 계산해 냈으며 이 상관행렬은 <표 4>와 같다.

일반적인 MTS 절차에 따라 정상집단으로부터 상관행렬과 역행렬 함수를 얻을 수 있고 이를 바탕으로 각각의 데이터에 대한 MD의 값을 계산 할 수 있다 31개 데이터의 MD의 평균값을 계산한 결과 평균 MD = 0.9696의 값을 얻었다. 이는 정상 집단의 MD 값은 1에 근접하므로 적절한 기준집단 선정이 이루어 졌다고 판단 할 수 있다. 또한 정상그룹을 MS를 기준으로한 파킨슨 환자의 MD 값을 계산한 결과 그 평균값은 412 로 상당히 큰 값을 나타내고 있다. 따라서 설정된 기준집단은 유효한 것으로 판단된다. <그림 1>은 정상인 그룹에서 설정된 MS를 기준으로 도출된 파킨슨환자 그룹의 MD 값을 도식화 하여 비교한 그래프 이다.

이러한 모든 계산과정은 Matlab 소프트웨어와 Excel을 이용하여 프로그램을 통해 계산 했다. 정상그룹의 MD 값은 2를 넘지 않았으며 파킨슨 환자의 MD 값은 정상 그룹과 비교하여 상당히 큰 값을 나타냈다. 따라서 선택된 정상그룹을 통한 MS는 효과적으로 파킨슨 환자의 진단에 활용될 수 있었다.

정상그룹을 기준으로 설정된 MS를 이용하여 이상치를 제거하기 전의 195개 전 데이터에 대한 진단을 실시하였다. 정상그룹과 파킨슨병을 구별하는 Thresh value는 2로 설정하였다. 진단된 결과는 정상인을 파킨슨병 환자로 진단한 Type 1 error와 파킨슨병 환자를 정

상으로 진단하는 Type 2 error의 수 그리고 정확히 진단한 진단율은 다음과 같다.

이상치 제거 전 Type 1 error로 나타난 42번 49번 50번 환자의 경우 8개의 데이터가 이상치로 나타났으며

따라서 이상치를 제거 할 경우 Type 1 error의 수는 15개에서 7개로 줄일 수 있었다. Type 2 error의 수는 이상치 제거 전후 3개로 변함이 없었다. <표 6>에서 42번, 49번, 50번 환자의 데이터 중 Type 1 error로 나타

<표 2> 정상그룹-정상인의 음성데이터

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16
237.3	243.7	229.3	0.003	0.00001	0.0017	0.0016	0.0052	0.012	0.116	0.007	0.007	0.0088	0.021	0.0053	24.68
243.4	250.9	232.4	0.002	0.000009	0.0011	0.0014	0.0033	0.014	0.126	0.008	0.009	0.0103	0.023	0.0045	25.37
198.4	215.2	193.1	0.002	0.00001	0.0011	0.0014	0.0034	0.013	0.111	0.006	0.008	0.0095	0.019	0.0012	30.78
242.9	255	227.9	0.002	0.000009	0.0012	0.0014	0.0035	0.015	0.134	0.008	0.009	0.0101	0.025	0.0048	25.03
229.4	252.2	221.2	0.002	0.000009	0.0011	0.0011	0.0034	0.015	0.129	0.008	0.01	0.0102	0.023	0.003	26.42
199.2	209.5	192.1	0.002	0.00001	0.0013	0.0014	0.004	0.01	0.089	0.005	0.006	0.0076	0.015	0.0017	30.94
237.2	247.3	225.2	0.003	0.00001	0.0017	0.0018	0.0051	0.018	0.164	0.01	0.01	0.0113	0.031	0.0074	22.74
197.1	206.9	192.1	0.003	0.00001	0.0017	0.0017	0.005	0.011	0.097	0.006	0.007	0.008	0.017	0.0034	26.78
222.2	231.3	205.5	0.003	0.00001	0.0015	0.0014	0.0046	0.016	0.145	0.009	0.011	0.012	0.026	0.0034	25.86
113.7	116.4	96.91	0.003	0.00003	0.0017	0.002	0.0051	0.015	0.133	0.007	0.009	0.0115	0.022	0.0048	26.55
229	239.5	113.2	0.002	0.00001	0.0014	0.0014	0.0041	0.017	0.154	0.009	0.012	0.0123	0.028	0.0042	24.55
228.8	234.6	223.6	0.003	0.00001	0.0018	0.0016	0.0053	0.016	0.145	0.009	0.011	0.0118	0.026	0.0035	25.96
245.5	262.1	231.8	0.002	0.00001	0.0013	0.0015	0.0038	0.016	0.141	0.009	0.01	0.0115	0.027	0.0048	24.6
202.3	211.6	197.1	0.002	0.000009	0.0009	0.0011	0.0028	0.01	0.085	0.005	0.006	0.0072	0.014	0.0007	32.68
126.3	134.2	112.8	0.004	0.00004	0.0013	0.0017	0.0039	0.02	0.185	0.011	0.01	0.0161	0.034	0.0047	25.03
260.1	264.9	237.3	0.003	0.00001	0.0021	0.0019	0.0062	0.02	0.197	0.012	0.012	0.0137	0.036	0.0091	21.08
252.5	261.5	182.8	0.002	0.000007	0.0009	0.0011	0.0028	0.012	0.103	0.006	0.007	0.0086	0.018	0.0043	26.81
122.2	128.6	115.8	0.005	0.00004	0.0017	0.002	0.0051	0.016	0.143	0.009	0.008	0.0143	0.026	0.0084	3.16
112.5	133.4	105.7	0.004	0.00003	0.0017	0.0019	0.005	0.014	0.129	0.007	0.008	0.0114	0.02	0.0044	26.44
241.4	248.8	232.5	0.003	0.00001	0.0016	0.0017	0.0047	0.018	0.154	0.01	0.01	0.0125	0.03	0.0068	23.15
201.5	210.6	195.7	0.002	0.00001	0.0011	0.0012	0.0031	0.012	0.107	0.006	0.008	0.0096	0.018	0.0014	31.73
128	138.1	122.1	0.004	0.00003	0.0014	0.0017	0.0041	0.023	0.21	0.013	0.011	0.0168	0.04	0.0048	24.69
115.4	123.1	108.6	0.003	0.00003	0.0016	0.002	0.0048	0.015	0.137	0.008	0.009	0.0113	0.024	0.004	26.01
116.4	129	109	0.003	0.00003	0.0017	0.0021	0.0051	0.017	0.155	0.009	0.01	0.0133	0.026	0.0042	26.14
124.4	135.1	117.5	0.004	0.00003	0.0014	0.0017	0.0042	0.022	0.197	0.012	0.01	0.0169	0.037	0.0048	25.14
214.3	260.3	77.97	0.006	0.00003	0.003	0.0032	0.0089	0.019	0.19	0.01	0.012	0.0137	0.031	0.044	21.21
129.3	139.9	118.6	0.005	0.00004	0.0017	0.0018	0.005	0.025	0.228	0.014	0.012	0.0195	0.042	0.0048	25.43
114.6	119.2	86.65	0.003	0.00003	0.0015	0.0018	0.0044	0.012	0.106	0.006	0.007	0.011	0.017	0.007	24.78
209.1	237.5	109.4	0.003	0.00001	0.0015	0.0015	0.0044	0.019	0.17	0.01	0.013	0.0138	0.029	0.0087	25.55
117	144.5	99.92	0.004	0.00003	0.0018	0.0022	0.0053	0.017	0.145	0.009	0.011	0.0132	0.026	0.0059	25.45
123	130	114.7	0.004	0.00003	0.0012	0.0016	0.0037	0.017	0.154	0.009	0.008	0.014	0.028	0.0046	24.97
209.5	253	89.49	0.006	0.00003	0.0033	0.0029	0.0099	0.028	0.263	0.016	0.017	0.0188	0.048	0.0181	19.15
110.7	113.6	100.1	0.004	0.00003	0.0017	0.002	0.0051	0.015	0.133	0.008	0.01	0.0129	0.023	0.0043	26.55
236.2	244.7	102.1	0.003	0.00001	0.0015	0.0015	0.0046	0.024	0.217	0.014	0.014	0.0162	0.042	0.0062	24.08
245	272.2	239.2	0.005	0.00002	0.0028	0.0024	0.0084	0.019	0.181	0.011	0.011	0.0126	0.033	0.0105	21.53
203.2	211.5	196.2	0.002	0.000009	0.0009	0.0011	0.0028	0.01	0.085	0.005	0.006	0.0073	0.014	0.0007	33.05
117.2	123.9	106.7	0.004	0.00004	0.0019	0.0027	0.0056	0.019	0.171	0.009	0.012	0.0195	0.026	0.0096	23.08
201.8	262.7	78.23	0.007	0.00003	0.0041	0.004	0.0124	0.026	0.255	0.015	0.016	0.0176	0.044	0.0444	19.37

<표 3> 비정상그룹-파킨슨 환자의 음성데이터

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16
100.8	115.7	85.55	0.01	0.0001	0.0062	0.0058	0.0187	0.031	0.361	0.018	0.018	0.0214	0.053	0.0222	19.2
169.8	191.8	151.5	0.016	0.00009	0.0086	0.0095	0.0259	0.081	0.821	0.038	0.054	0.0881	0.114	0.0753	12.36
143.5	162.2	65.81	0.011	0.00008	0.0065	0.0047	0.0194	0.054	0.478	0.032	0.024	0.0339	0.095	0.0488	20.34
151.9	492.9	69.09	0.009	0.00006	0.004	0.0039	0.0121	0.018	0.235	0.008	0.008	0.0127	0.024	0.0605	23.69
120.6	131.2	113.8	0.01	0.00008	0.0046	0.0075	0.0139	0.047	0.456	0.023	0.035	0.0324	0.07	0.0122	21.38
114.8	271.3	104.7	0.009	0.00008	0.0037	0.0052	0.0112	0.032	0.35	0.018	0.02	0.0252	0.054	0.0114	21.66
116	141.8	110.7	0.013	0.00011	0.0066	0.0091	0.0197	0.064	0.584	0.035	0.048	0.0447	0.105	0.0177	19.65
148.1	162.8	67.34	0.008	0.00005	0.0047	0.0035	0.014	0.054	0.497	0.034	0.025	0.0364	0.101	0.0243	21.72
152	163.3	147.2	0.004	0.00003	0.0022	0.0023	0.0067	0.08	0.772	0.044	0.05	0.0569	0.133	0.0166	19.66
114.2	124.4	77.02	0.006	0.00005	0.003	0.0032	0.009	0.04	0.406	0.016	0.02	0.0411	0.047	0.0207	20.44
152.1	161.5	76.6	0.004	0.00003	0.0019	0.0023	0.0057	0.059	0.637	0.033	0.038	0.044	0.1	0.0121	20.97
202.6	565.7	177.3	0.016	0.00008	0.0092	0.0096	0.0276	0.072	0.833	0.035	0.043	0.0646	0.105	0.0789	14.99
186.7	212	149.4	0.019	0.0001	0.0108	0.0115	0.0323	0.058	0.784	0.027	0.037	0.0626	0.081	0.1095	12.53

<표 4> 상관행렬

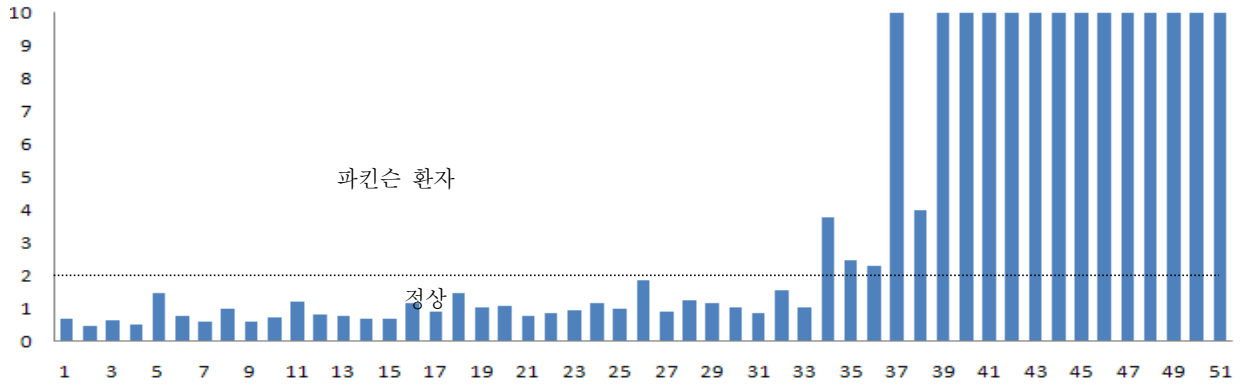
	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16
A1	1	0.984	0.78	-0.52	-0.86	-0.03	-0.32	-0.03	-0.14	-0.12	-0.06	0.121	-0.4	-0.06	0.112	-0.09
A2	0.984	1	0.692	-0.43	-0.81	0.078	-0.21	0.077	-0.07	-0.04	0.001	0.212	-0.33	9E-04	0.234	-0.15
A3	0.78	0.692	1	-0.67	-0.82	-0.38	-0.6	-0.38	-0.36	-0.37	-0.27	-0.25	-0.55	-0.27	-0.34	0.198
A4	-0.52	-0.43	-0.67	1	0.835	0.712	0.807	0.712	0.708	0.739	0.676	0.476	0.807	0.676	0.59	-0.62
A5	-0.86	-0.81	-0.82	0.835	1	0.367	0.616	0.367	0.478	0.479	0.415	0.189	0.692	0.415	0.254	-0.3
A6	-0.03	0.078	-0.38	0.712	0.367	1	0.915	1	0.55	0.615	0.536	0.657	0.484	0.536	0.756	-0.69
A7	-0.32	-0.21	-0.6	0.807	0.616	0.915	1	0.914	0.518	0.575	0.486	0.535	0.537	0.486	0.771	-0.65
A8	-0.03	0.077	-0.38	0.712	0.367	1	0.914	1	0.55	0.615	0.535	0.658	0.484	0.535	0.755	-0.69
A9	-0.14	-0.07	-0.36	0.708	0.478	0.55	0.518	0.55	1	0.992	0.991	0.87	0.948	0.991	0.365	-0.69
A10	-0.12	-0.04	-0.37	0.739	0.479	0.615	0.575	0.615	0.992	1	0.986	0.873	0.932	0.986	0.446	-0.71
A11	-0.06	0.001	-0.27	0.676	0.415	0.536	0.486	0.535	0.991	0.986	1	0.854	0.913	1	0.358	-0.71
A12	0.121	0.212	-0.25	0.476	0.189	0.657	0.535	0.658	0.87	0.873	0.854	1	0.743	0.854	0.437	-0.68
A13	-0.4	-0.33	-0.55	0.807	0.692	0.484	0.537	0.484	0.948	0.932	0.913	0.743	1	0.913	0.317	-0.61
A14	-0.06	9E-04	-0.27	0.676	0.415	0.536	0.486	0.535	0.991	0.986	1	0.854	0.913	1	0.358	-0.71
A15	0.112	0.234	-0.34	0.59	0.254	0.756	0.771	0.755	0.365	0.446	0.358	0.437	0.317	0.358	1	-0.58
A16	-0.09	-0.15	0.198	-0.62	-0.3	-0.69	-0.65	-0.69	-0.69	-0.71	-0.71	-0.68	-0.61	-0.71	-0.58	1

난 경우는 대부분 이상치로 나타났다. 특히 이상치를 포함하고 있는 환자의 진단에러가 많이 발생하여 이는 환자 개개인의 음성 특성 또는 측정상의 데이터 측정상의 문제로 판단된다. Type 1 error는 파킨슨병이 아닌 환자를 파킨슨병으로 진단하는 경우이므로 Type 2 error에 비해 진단오류로 인한 과급 효과는 크지 않다. Type 2 error 즉 파킨슨병 환자를 정상으로 진단하는 error는 이상치 제거전후 8번 환자의 3번, 5번 데이터에서 동시

에 발생하였으나 나머지 4번의 측정값은 정확한 진단이 이루어 졌다. 44번과 19번 환자의 경우 6번의 측정 데이터 중 1번의 Type 2 error가 발생하였다.

3.3 유효 변수의 선정

마지막으로 직교 L32 직교 배열표와 망대특성의 SN비를 이용하여 각각의 변수의 중요도를 분석하였다. 16



<그림 1> 정상집단과 비정상집단의 MD 값의 비교

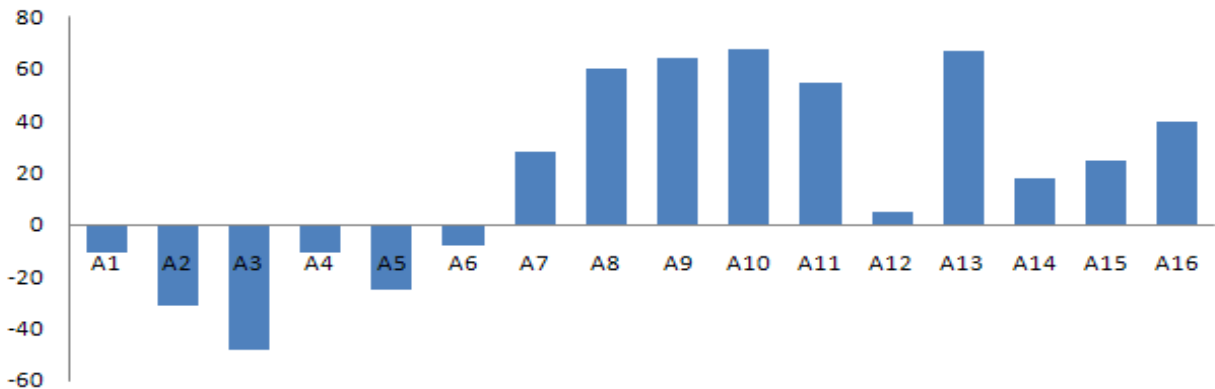
<표 5> 파킨슨 환자의 진단 결과

	이상치 제거 전	이상치 제거 후
Type 1 error	15	7
Type 2 error	3	3
진단 성공률	90.77%	94.48%

개의 변수를 적절하게 배치하기 위해서는 L32 OA를 사용하였으며 망대특성의 SN비를 사용하였다. 식 (3) 직교 배열표에서 Level 1은 변수를 사용한 경우이고 Level 2는 변수를 사용하지 않은 경우를 의미 한다. 직교배열표의 조합에 따라 13개의 비정상 집단의 데이터를 이용하여 MD 값을 계산하였으며 이를 근거로 각각의 SN비를 계산 하였다.

<표 6> 이상치 (outlier)와 Type 1 and Type 2 error

error 종류	환자번호	이상치 제거 전	환자 번호	이상치 제거 후
Type1-error	01	1 2 3 4 ⑤ 6	01	1 2 3 4 ⑤ 6
	17	1 ② 3 4 5 6	17	1 ② 3 4 5 6
	42	① 2 3 ④ ⑤ ⑥	42	① 2 3 ④ ⑤ ⑥
	49	① ② ③ ④ ⑤ 6	49	① ② ③ ④ ⑤ 6
	50	① ② 3 ④ ⑤ 6	50	① ② 3 ④ ⑤ 6
Type2-error	08	1 2 ③ 4 ⑤ 6	08	1 2 ③ 4 ⑤ 6
	44	1 2 ③ 4 5 6	19	① 2 3 4 5 6



<그림 2> 변수들의 Gain 값

<표 7> 최적화 전후의 진단 결과의 변화

	16개 변수를 이용한 진단	13개 변수를 이용한 진단
Type 1 error	7	8
Type 2 error	3	3
진단 성공율	94.48%	93.92%

$$Gain = SN\text{비 해당변수를 사용한 경우} \quad (4)$$

$$SN\text{비 해당 변수를 사용하지 않은 경우}$$

식 (4)에서 양의 Gain 값이 클수록 변수는 마할라노비스 거리(MD)에 긍정적인 효과를 미치는 것으로 판단된다. 따라서 큰 Gain 값을 가지는 변수는 파킨슨병에 매우 긍정적인 영향을 미치는 것으로 볼 수 있으며 음의 Gain 값을 가지는 변수는 파킨슨병의 진단에 큰 영향을 미치지 못하는 변수로 판단할 수 있다. 따라서 파킨슨병의 진단을 위한 데이터를 확보하기 위한 시간적 경제적 측면에서 무시될 수 있을 것이다. <그림 2>는 변수에 따른 Gain 값을 나타내고 있다.

다음은 상대적으로 큰 음의 Gain 값을 가지는 변수 A2, A3, A5를 변수 항목에서 제거한 후 13개의 변수만을 가지고 파킨슨병의 진단을 실시하여 전 변수를 사용하여 진단한 결과와 비교하여 보았다<표 7>.

4. 결 론

본 연구에서와 같이 다수의 변수를 가지는 다변량 시스템의 정확한 해석을 위해서는 많은 양의 데이터와 전문적인 지식이 요구된다. 또한 다변량 해석은 필요한 데이터를 얻기 위한 시간적인 경제적인 제한 때문에 일부 제한된 데이터를 효과적으로 분석하는 방법에 대한 많은 연구가 이루어져 왔다. 그동안 진행된 대부분의 연구는 수학 및 통계적 접근 방법으로 매우 복잡한 과정과 절차 및 통계학적인 전문지식을 필요로 한다.

MTS 기법은 새로운 데이터 해석방법(Data Analytic Method)으로 암과 같은 질병의 진단, 문자 인식, 음성인식, 최적화 등의 분야에서 활발한 연구가 이루어지고 있고 성공적인 적용사례도 여러 논문을 통해 발표되고 있다. 진단을 위한 MTS 기법은 적절한 정상집단 및 비정상 집단의 선정을 통한 Mahalanobis Space만 만들 수 있다면 진단에 필요한 전문적인 지식 없어도 효과적으로 활용할 수 있다. 본 연구에 사용된 데이터는 옥스퍼드 대학에서 Nonlinear Recurrence를 통한 해석에 사

용된 동일 데이터를 사용하였다. 이 방법을 통한 진단 성공률은 $91.8 \pm 2.0\%$ 로 이상치를 제거한 후 얻은 본 연구의 진단 성공률 94.48%보다 낮은 것으로 나타났다. 또한 변수가 진단에 미치는 영향을 분석하여 불필요한 변수를 제거함으로써 진단의 위한 시간적 경제적으로 효과적인 방법이 될 수 있을 것으로 판단된다.

참고문헌

- [1] 장영애, 김용덕, 반제천, 성철재; “파킨슨병 환자와 정상노인의 음성비교”, 말소리와 음성과학, 1(1) : 99-107, 2009.
- [2] 이옥분, 정옥란, 고도홍; “강도 및 음도조절을 이용한 훈련이 파킨슨병 환자의 음성 및 발화 명료도 개선에 미치는 효과”, 음성과학 8(3) : 173-184, 2001.
- [3] 최성희, 김향희, 이원용, 최홍식; “파킨슨증의 음성진전 : 감별진단을 위한 예비연구” 음성과학, 12(3) : 19-23, 2005, 9.
- [4] Max A Little, P. E. Mcshary, S. J. Roberts, D. A. Costello and I. M. Moroz; “Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection,” Biomedical Online, 6 : 23, 2007, 6.
- [5] Felix Javier Jimenez-Jimenez, et. al.; “Acoustic Voice analysis in untreated patients with Parkinson’s disease,” Parkinson’s and Related Disorders, 3(2) : 111-116, 1997, 4.
- [6] Julia K. MacCallum, et al.; “Acoustic Analysis of aperiodic voice perturbation and nonlinear dynamic properties in esophageal phonation,” Journal of Voice, 23(3) : 283-290, 2003, 5.
- [7] Javier Gamboa et al.; “Acoustic Voice Analysis in patients with Parkinson’s disease treated with dopaminergic drugs,” Journal of Voice, 11(3) : 314-320, 1997, 9.
- [8] Taguchi, G and R. Jugulum; The Mahalanobis-Taguchi Strategy : A Pattern Technology System, John Wiley and Sons, Inc., 2002.
- [9] Taguchi, G and R. Jugulum; “New Trends in Multivariate Diagnosis,” Indian Journal of Statistics, 62(B) : 2 233-248, 2000.
- [10] 홍정의, 권홍규; “Mahalanobis Taguchi System을 이용한 다변량 시스템의 해석에 관한 연구”, 한국 산업 경영시스템학회지, 32(1) : 20-25, 2009, 3.
- [11] 홍정의, 권홍규; “Mahalanobis Taguchi System을 이용한 자동차 브레이크 성능 만족도를 고려한 설계조건 선정에 관한 연구”, 한국 산업경영시스템 학회지, 30(1) : 41-47, 2007, 3.
- [12] Hong, Jung-Eui, Cudney E. A., Taguchi G, Jugulum R.,

- Paryani K. and Ragsdell K.; "A Comparison study of Mahalanobis Taguchi System and neural network for multivariate pattern recognition," 2005 ASME IMECE Proceedings, Orlando FL.
- [13] E. A. Cudney, Jung-Eui Hong, Rajesh Jugulum, Kioumars Paryani, Kenneth M. Ragsdell and Genishi Taguchi; "An Evaluation of Mahalanobis Taguchi System and Neural Network for Multivariate Pattern Recognition," *Journal of Industrial and System Engineering*, 1(2) : 139-150, 2007.
- [14] Lande, U.; "Mahalanobis Distance : A Theoretical and Practical Approach," <http://biologi.uio.no/fellesavdelinger/finse/spatialstats/Mahalanobis%20distance.ppt>, 2003.
- [15] Hayashi, S., Y. Tanaka, and E. Kodama; "A New Manufacturing Control System using Mahalanobis Distance for Maximizing Productivity," *IEEE Transactions*, 59-62, 2001.
- [16] R. Jugulum, G. Taguchi, S. Taguchi and J. Wilkins; "Discussion of A review and analysis of Mahalanobis- Taguchi System," *Technometrics*, 45(1), 2003, 2.
- [17] R. Jugulum and L. Monplaisir; "Comparison between Mahalanobis-Taguchi-System and Artificial Neural Networks," *Journal of Quality Engineering Society*, 10(1) : 60-73, 2002.