

웹 기반 프랑스어 음소 지각 훈련과 그 신뢰도 평가

김 선 희¹ · 류 나 영² · 원 용 국^{3*}

¹서울대학교 사범대학 불어교육과 조교수

²펜실베이니아주립대학교 아시아학과 조교수

^{3*}서울대학교 교육종합연구원 객원연구원

Reliability of Web-based French Phoneme Perception Test

Sunhee Kim¹ · Na-Young Ryu² · Yongkook Won^{3*}

¹Assistant Professor, Department of French Language Education, Seoul National University, Seoul 08826, Korea

²Assistant Teaching Professor, Department of Asian Studies, Pennsylvania State University, PA 16803, USA

^{3*}Visiting Researcher, Center for Educational Research, Seoul National University, Seoul 08826, Korea

[요 약]

본 연구는 프랑스어를 배우는 한국인 학습자를 대상으로 설계한 웹 기반의 음소 지각 훈련 시스템을 소개하고 이 시스템의 신뢰도를 평가하였다. 제안한 지각 훈련은 자음 및 모음에 대한 분별 테스트와 식별 테스트의 총 4개의 테스트로 구성되었다. 그 신뢰도를 분석하기 위하여 먼저 지각 훈련 결과의 기술 통계량을 살펴보고, 준거참조평가 방법으로 지각 훈련의 신뢰도를 분석하였다. 지각 훈련의 내적일관성과 신뢰도는 크론바흐 알파 (Cronbach's α) 값과 의존도 계수 (dependability Φ)를 활용하였다. 제안한 4개의 테스트에 해당하는 문항 전체를 하나로 묶었을 때 크론바흐 알파는 0.71~0.97로 대체로 높게 나타났고, 의존도 계수 Φ 는 0.65~0.97로 나타났다. 문항 수가 적은 일부 하위 테스트의 의존도 계수가 낮게 나타나는데, 이러한 경우는 신뢰도를 높일 수 있도록 다른 항목과 묶어서 하나의 테스트로 활용하는 방법을 제안하였다.

[Abstract]

This study introduces and assesses the reliability of a web-based phoneme perception training system created for Korean learners studying French. The suggested perception training test consisted of four subtests: consonant and vowel discrimination tests, and consonant and vowel identification tests. To assess the reliability of the perception training test, descriptive statistics of perception training results were reviewed first, followed by a criterion-referenced evaluation approach. The internal consistency and reliability of the perceptual training test were assessed using Cronbach's alpha and dependability phi. The perception test had a Cronbach's alpha of 0.71 to 0.97, which was generally high, and a dependability coefficient of 0.65 to 0.97. Low dependability ratings were found in a few subtests with a small number of items. It is suggested that these subtests be put together as one test to increase the test's reliability.

색인어 : 신뢰도 평가, 웹 기반 테스트, 음소, 의존도 계수, 지각훈련

Keyword : Dependability, Perception training, Phoneme, Reliability, Web-based test

<http://dx.doi.org/10.9728/dcs.2021.22.12.2099>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 19 November 2021; **Revised** 07 December 2021

Accepted 07 December 2021

***Corresponding Author, Yongkook Won**

Tel: +82-2-880-5721

E-mail: linguistry@gmail.com

I. 서론

이전의 외국어 학습이 주로 문법을 중심으로 한 읽기 능력과 쓰기 능력을 향상시키는 데 중점을 두었다면 최근에는 외국어를 실생활에서 이용하는 것에 대한 요구가 일반화되면서 말하기 능력과 듣기 능력에 대한 관심이 증대되었다. 영어의 경우에 국내에서 최근 성인을 대상으로 한 학습 시장이 확대되었고 이러한 경향은 최근의 정보기술 및 인공지능 기술을 기반으로 한 온라인 교육과 함께 뚜렷하게 나타난다. 국외의 경우에도 인터넷의 보급과 함께 공용어로서의 영어 사용이 확대되고 있고, 유럽의 경우는 유럽 연합의 출발과 함께 의사소통 중심 이론이 좀 더 심화된 과제 중심 이론으로 바뀌면서 실생활에서의 의사소통에 대한 필요가 좀 더 구체화되었다고 볼 수 있다[1]. 이와 같이 의사소통 중심 혹은 과제 중심의 외국어 학습에 있어서 특히 말하기와 듣기는 어느 때보다 그 학습에 대한 중요성이 부각되고 있으나 실제로 효과적인 학습을 수행하는 것은 쉽지 않다.

외국어 발음 습득에 있어서 어려움에 대한 설명을 제시한 이론으로는 Flege의 음성학습 모델 (Speech Learning Model)[2]과 Best의 지각동화 모델 (Perceptual Assimilation Model)[3],[4]이 있다. 이 두 모델은 외국어 학습자들의 경우에 모국어인 L1의 전이가 L2의 습득에 영향을 미친다는 것을 다수의 후속 연구들을 통하여 규명하였다. 음성학습 모델과 지각동화 모델은 모두 L1 전이와 L2 경험을 제 2 언어 음성/음운 습득의 중요한 요소로 보고 있으며, 이 모델을 기반으로 이후 다양한 L2 지각 훈련 시스템을 제시하고 있다. 즉, 외국어 학습자들의 발음 교육에 있어서 L1과 L2에 대한 기본적인 차이를 근거로 발음의 생성과 지각 훈련이 이루어질 필요가 있다는 것이다.

말하기와 듣기에 있어서 기본이 되는 음소에 대한 교육은 실제로 교실 환경에서 이루어지는 데에는 여러 가지 제약이 있다. 특히 국내뿐만 아니라 국외에서도 영어에 비하여 그 수요가 크지 않은 프랑스어의 경우는 한국인 학습자를 대상으로 하는 정보기술이나 인공지능 기술을 기반으로 한 학습 앱도 많지 않고, 기본적인 음소 교육을 위한 훈련이나 학습 시스템은 더욱 부족한 실정이다[5].

본 연구는 기본적으로 프랑스어를 배우는 한국인 학습자를 대상으로 L1인 한국어와 L2인 프랑스어의 음소 체계의 차이를 근거로 하여 설계한 대립쌍을 이용하여 개발한 웹 기반 전체 음소의 지각 훈련 시스템을 소개하고 이 시스템의 신뢰도를 평가하는 것을 그 목적으로 한다. 이러한 연구는 프랑스어 듣기와 말하기 학습에 기본이 되는 음소 습득 훈련용 시스템에 대한 신뢰도를 검증할 수 있고, 검증한 시스템 신뢰도를 바탕으로 해당 시스템을 중등학교 및 대학에서 제2외국어로서 프랑스어를 배우는 학습자들에게 활용할 수 있을 것이다. 또한, 현재는 기본적으로 두 언어 체계의 차이를 바탕으로 한 모든 대립쌍을 제공하고 있으나 이후 학습자들의 데이터 분석을 기반으로 하여 학습자 맞춤형의 지능적인 시스템으로

더욱 발전시킬 수 있을 것으로 본다.

본 논문의 구성은 다음과 같다. 먼저, 2장에서는 관련 연구들을 소개하고, 3장에서는 웹 기반 프랑스어 지각 실험 방법을 제시한다. 4장에서는 실험 결과에 이에 대한 평가 결과를 제시하고, 5장의 결론으로 마무리한다.

II. 관련 연구

2-1 제2언어 음성/음운 습득 모델

1) 음성학습 모델 (Speech Learning Model)

음성학습 모델은 성인 학습자가 외국어 발음과 지각의 어려움 겪는 이유를 설명하는 모델이다[2],[6],[7]. 이 모델에서는 학습자가 모국어 (L1)와 목표 언어 (L2)의 음소를 어떻게 대응시켜 지각하는지를 3가지 패턴으로 보여준다. L1과 L2 음의 일대일 대응을 통해, L2를 ‘동일한 (identical)’, ‘새로운 (new)’, ‘유사한 (similar)’ 음으로 분류한다. 먼저, 학습자가 L1과 L2 음을 동일한 소리로 인지한다면, L1의 긍정적 전이로 인해 L2 음 습득을 쉽게 할 수 있게 된다.

예를 들어서 한국어 /i/와 일본어의 /i/가 음성학적으로나 음향적으로도 차이가 없어서, 일본인 화자는 한국어 /i/를 일본어 /i/와 동일하다고 인지하고, 한국어 /i/를 습득하는 데 어려움이 없다. 학습자에게 L2 음소가 새로운 소리로 인식되면, L1의 간섭이 거의 없는 것으로 간주되고, L2 음에 충분히 노출된다면 이 또한 학습자가 어려움 없이 새로운 L2 음을 습득할 수 있다. 반면, L2 음이 L1 음과 음향이나 지각적으로 유사하게 인식되는 경우에는 습득에 어려움이 있다. 예를 들면, 한국인 영어 학습자의 경우, 영어의 두 음소 /e/와 /æ/가 한국어 /e/와 유사하게 인식되어서, 두 음소를 정확하게 발화하거나 구별하는 데 상당한 어려움을 겪는다.

이처럼 음성학습 모델은 L1 전이가 L2 음 습득에 영향을 미치는 중요한 요소이며, 학습자의 L2 경험이 늘수록, L2 지각 (perception)과 발화 (production) 능력이 향상된다고 주장한다.

2) 지각동화 모델 (Perceptual Assimilation Model)

지각동화 모델[3],[8]은 음성학습 모델과 같이 L1 전이를 논의의 출발점으로 한다는 점에서 유사하나, 외국어를 배운 경험이 없는 학습자가 L1과 L2 음소 사이에서 지각된 음성학적 유사성을 바탕으로 어떻게 동화되는지를 중점적으로 본다. 이러한 지각동화 모델을 L2 연구에 확장시킨 모델로는 L2 지각동화 모델 (PAM-L2)[4]이 있다.

PAM 모델에서는 L1과 L2 사이의 지각 동화 정도에 따라서, 두 범주 동화 (Two Category Assimilation), 범주 적합도 (Category Goodness), 단일 범주 동화 (Single Category Assimilation), 범주화 불가능 (Uncategorized)으로 분류하며, 이를 통해서 학습자가 L2 음을 얼마나 잘 지

각할 수 있는지를 예측한다. ‘두 범주 동화’는 L2의 서로 다른 음소가 L1의 각기 다른 음소로 동화되는 것으로 L2 범주가 정확하게 인식되어서 습득에 어려움이 없다. 또한, ‘범주화 불가능’ 분류처럼 L2에 동화되는 L1이 없는 경우도 L2 경험과 관계없이 L2 범주가 정확하게 인식되어서 습득에 어려움이 없다. 하지만, ‘단일 범주 동화’는 서로 다른 두 개의 L2 음이 L1에서는 하나의 범주에 존재하는 소리로 대응하기 때문에 습득에 어려움이 있다. 예를 들어서, 한국인 영어 학습자는 영어의 두 다른 음소인, 긴장 모음 /i/와 이완 모음 /i/가 한국어 모음 /i/ 하나의 범주로 동화되기 때문에, 두 음소를 정확하게 구별해서 지각하기 쉽지 않다.

2-2 지각 훈련 방법론

그동안 많은 연구자가 학습자의 L2 지각과 발음 향상에 기여하기 위해서 다양한 지각 훈련 방법론을 제안해왔다. 특히 훈련 장소(실험실 vs. 온라인), 훈련 과제 유형(음소 식별 테스트 vs. 음소 분별 테스트), 훈련 음소 수(모든 L2 음소 vs. 어려운 L2 음소)에 따라서 훈련의 효과가 다를 수 있기 때문에, 본 논문에서는 이 부분에 대해 중점적으로 살펴보고, 효율적인 지각 프로그램 개발을 위한 방향을 제시해 보고자 한다.

1) 훈련 장소: 실험실 기반과 온라인 기반 훈련 (Lab-based training vs. Online training)

지각 훈련 연구에서 고려해야 할 요소 중의 하나는 학습자가 어디에서 훈련을 받느냐는 것이다. 기존의 대다수 지각 훈련은 실험실에서 진행이 되었으며, 학습자는 이러한 훈련을 통해 L2 지각 및 발음 향상을 보여주었다. 하지만, 실험실 기반 훈련은 일반적으로 정해진 일정에 따라 수행되어야 하며, 학습자 중심적이고 장기적인 접근 방식의 이점을 가지기 어렵다. 이 때문에 온라인 훈련에 대한 많은 관심과 개발이 이루어지고 있다. 예를 들어, Sakai와 Moorman은 지각훈련 연구에 관한 메타 분석을 통해 집에서 참여한 훈련이 실험실보다 더 효과적이었다는 것을 보여주었으며[9], Thomson과 Derwing은 온라인 지각 프로그램인 English Accent Coach[10]를 이용하여 영어 모음 발음 훈련을 수행하여 효과가 있음을 입증하였다[11]. 이뿐만 아니라, Motohashi-Saigo와 Hardison은 일본어를 배우는 영어 원어민 화자를 위한 온라인 시청각 발음 훈련 프로그램을 개발했으며, 학습자가 플랫폼의 유연성과 훈련 중에 받은 즉각적인 피드백에 만족한다는 설문 결과를 보여주었다[12].

이처럼, 웹을 기반으로 한 지각 및 발음 훈련 프로그램의 효과뿐만 아니라 참여자의 편의성과 실험자의 대응량 데이터를 수집할 수 있는 장점 때문에, 온라인 언어 실험 및 교육 프로그램이 많은 관심을 받고 있다. 즉, 최근 실험실 기반과 온라인 기반의 연구를 비교하는 기술적인 분석이 많은 주목을 받고 있다. Bridges et al.은 실험실에서 주로 사용되는 PsychoPy, OpenSesame, E-Prime 등의 프로그램과 온라인

에서 사용되는 jsPsych, Gorilla, PshcyoJS 프로그램의 성능과 반응 속도에 따른 데이터 정확도를 비교하는 연구를 하였다[13]. 분석 결과 실험실 기반 연구에서 사용된 프로그램의 성능과 정확도가 높은 것으로 나타났으며, 온라인 실험에서 이용되는 프로그램의 경우 각기 다른 사용자의 시스템과 브라우저가 실험에 영향을 줄 수 있음에도 불구하고, 정교한 반응 속도 연구를 제외한 실험에서는 유용하게 사용될 수 있다.

앞으로 L2 교육 및 훈련 프로그램뿐만 아니라 언어 실험을 하는 데 있어서, 웹이나 앱을 활용하는 것이 참여자에게는 자율성과 편리성을 제공하고, 연구자에게는 단기간에 대량의 데이터를 수집할 수 있는 효율적인 방법이 될 수 있을 것으로 전망한다.

2) 훈련 과제: 음소 분별 테스트와 음소 식별 테스트 (Discrimination task vs. Identification task)

대부분의 지각 훈련에서는 음소 분별 테스트 또는 음소 식별 테스트를 사용하거나 둘 다 활용하는 경우가 있는데, 어떤 훈련 테스트가 효율적인지 살펴볼 필요가 있다. 참고로, 음소 분별 테스트는 두 개의 음성 자극을 듣고 두 소리가 같은지 다른지를 판별하는 것이고, 음소 식별 테스트는 한 개의 음성 자극을 듣고 그 음소가 여러 발음 유형 중에서 어디에 속하는지를 파악하는 것이다.

기존 연구에서는 음소 식별 훈련이 음소 분별 훈련보다 더 효과적이라고 제안하고 있다[14]-[17]. 예를 들어, Lively et al.[16]과 Logan et al.[17]은 음소 식별 훈련을 받은 학습자들은 훈련되지 않은 자극에 대해 지각 학습을 일반화할 수 있다고 했다. 그리고, Strange와 Dittmann은 음소 분별 훈련을 받은 학습자들은 훈련되지 않은 자극에 대한 학습을 일반화하지 않았다는 것을 보여주었다[19]. 그러나 다른 연구자들은 식별 및 분별 훈련 방법이 함께 활용될 때, 지각 정확도를 향상시킨다는 것을 보여주었다[20],[21]. 최근 연구에서 Shinohara와 Iverson은 일본인 학습자를 대상으로 영어 /r-/l/을 위한 음소 식별 훈련과 음소 분별 훈련을 효과를 비교해보니, 두 훈련 방법을 통합할 때 지각 정확도가 가장 높다는 것을 보여주었다[20].

3) 훈련 음소 수: 모든 L2 음소 vs. 어려운 L2 음소 (The full set of phonemes vs. Difficult phonemes)

대부분의 연구에서는 특정 L2 모음과 자음만을 이용하여 지각 및 발음 훈련을 했으나, 실질적으로 학습자를 위한 훈련 및 테스트 프로그램을 개발할 때, 보다 폭넓은 L2 음소가 추가될 필요가 있다. 이에 관해서 Nishi와 Kewley-Port는 일본인 영어 학습자를 대상으로 영어 단모음 9개와 어려운 단모음 3개만을 이용해 지각 실험을 한 결과, 9개의 모든 모음을 훈련받은 학습자가 모든 모음에서 지각률이 상승하였고, 가장 어려운 3개의 모음만 집중한 학습자는 3개의 모음 외에 훈련 받지 않은 다른 모음에 대해서는 지각률 상승이 없었다[22]. 또한, Nishi와 Kewley-Port의 차후 연구에서 한국인 학습자

를 대상으로 연구를 진행했을 때도, 동일한 결과를 얻었다. 이 연구 결과를 통해서 학습자의 실력 향상을 위해 모든 L2 음소를 훈련하는 것이, 어렵다고 알려진 특정 음소에만 집중하는 것보다 더 도움이 될 것이라고 생각된다[23].

위의 지각 훈련에 관한 방법론에서 제시한 것을 바탕으로, 본 연구에서는 jsPsych[24]를 이용해서 웹 기반 프랑스어 훈련 프로그램을 개발했으며, 학습자는 이 프로그램 (음소 식별 테스트와 음소 분별 테스트)을 통해서 프랑스어의 모든 자음과 모음을 훈련할 수 있다.

2-3 L2 지각 훈련의 효과

지난 25년 동안 많은 연구자가 학습자들의 L2 분절음과 초분절음 지각 능력과 발음을 향상시키기 위해서 다양한 형태의 훈련 프로그램을 제시하였다[16],[22],[25]-[32].

대부분의 연구에서 학습자는 지각 훈련을 통해 L2 음을 정확하게 구별하고 인지하는 능력이 향상되었으며, 이 효과가 장기적으로도 유지되고 있음을 보여주고 있다[19]. Logan et al.은 처음으로 HVPT (High variability phonetic training) 방법을 제시하였는데, 이는 지각 훈련에서 다양한 L2 화자와 L2 음운 환경을 제공함으로써, 학습자가 L2 음의 음향적 신호를 감지하고, 식별하는 능력을 키울 수 있게 하는 것이다[18]. Lively et al.도 일본인 학습자에게 HVPT를 기반으로 한 영어 /r-/l/ 지각 훈련을 수행했는데, 훈련 후에 두 음소를 구별하는 능력이 상당히 향상되었다는 결과를 보여주었다[16]. 또한, Pierce 연구에서는 영어를 배우는 중국인과 한국인 학습자를 대상으로 영어 모음 지각 훈련을 시킨 결과, 영어 모음 지각 향상뿐만 아니라, 새로운 단어에서도 일반화하는 효과가 있다는 것을 보여주었다[33].

이 밖에도 지각 훈련이 L2 음소 구별 및 인지 능력 향상뿐만 아니라, L2 발음에도 영향을 준다는 몇몇 연구가 있다. 흥미롭게도, 지각 훈련이 L2 발음에 미치는 영향은 상반된 결과를 보여주고 있다[26],[28],[34]-[37]. Bradlow et al.은 성인 일본인 학습자에게 영어 /r-/l/ 지각 훈련을 시킨 후, 지각과 발음 향상에 얼마나 영향이 있는지를 조사했는데, 지각 훈련을 받은 일본인 학습자는 사후 테스트에서 두 영어 음소를 구별하는 지각 능력뿐만 아니라 발음도 향상되었음을 보여주었다[26]. 이뿐만 아니라, 이 성과가 3개월 후에도 유지된다는 것을 발견했다. 이는 발음 훈련을 따로 제공하지 않았음에도 불구하고, 지각 훈련만으로도 충분히 L2 발음을 향상시킬 수 있으며, 이는 지각 훈련 시스템이 성공적인 L2 학습 도구가 될 수 있음을 보여주는 것이다. 반대로, Nobre-Oliveira는 L2 지각 훈련이 모든 L2 음소의 발음 향상을 가져오지는 않는다는 증거를 보여주었다[35]. 이 연구에서는 브라질 학습자에게 영어 모음 3쌍의 /i/-/ɪ/, /e/-/æ/, /ʊ/-/u/ 음소 분별 훈련을 시켰으며, 지각 정확도는 3쌍 모두 현저하게 증가하였다. 하지만, 음향 분석을 통한 발음 효과를 보면 영어 /i/-/ɪ/에서는 발음 향상을 보여주었으나, /e/-/æ/

와 /ʊ/-/u/에서는 향상되지 않았다.

이처럼 지각 훈련이 확실히 L2 지각을 구별하고 인식하는 데에는 효율적인 학습 도구가 될 수 있다는 것을 알 수 있으나, L2 발음 향상에는 얼마나 영향을 미치는지에 대해서는 추가 연구가 필요해 보인다. 또한, Lee et al.의 L2 발음 메타 분석에 따르면 기존 연구의 97%가 영어를 연구 대상으로 하고 있는데[38], 영어뿐만 아니라 다양한 외국어에 대한 지각 훈련 연구가 추가되어야 하며, 효율적인 훈련 프로그램 개발을 위해서 여러 가지 방법적인 측면을 고려해 보아야 할 것이다.

2-4 언어평가의 신뢰도

언어평가는 일반적으로 참고 기준을 근거로 표준참조평가(norm-referenced test)와 준거참조평가(criterion-referenced test)의 두 가지로 분류된다[39].

표준참조평가란 수험자 전체의 성취수준이 표준정규분포를 이룬다고 가정하고, 수험자의 성적이 집단 내에서 상대적으로 어디에 위치하는지 파악하는 데 목표를 둔 일종의 상대평가에 해당한다. 이러한 표준참조평가는 외국어능력, 강의 청취 능력 등 추상적이고 넓은 개념의 언어 특질을 평가하는데 활용한다[40]. 표준참조평가에서 수험자의 성적은 집단 내에서 상위 또는 하위 몇 %에 속하는지를 보여준다.

반면에, 준거참조평가는 수험자가 특정 성취 과제를 얼마만큼 알고 있는지, 특정 언어기술이 어느 수준에 도달했는지 확인하는 데 사용한다. 준거참조평가는 평가기준 (scoring rubric)에 따라 초·중·고급 등으로 수험자의 수준을 나눈다[39],[40]. 본 연구의 대상이 되는 <웹 기반 프랑스어 지각 훈련>은 수험자의 성취 점수를 상대적 순서로 나열하는 게 아니라, 수험자가 프랑스어 발음을 제대로 지각하는지 평가하므로 준거참조평가에 해당한다.

언어평가의 질을 평가하는 평가 신뢰도는 언어평가의 종류에 따라 다르다. 표준참조평가의 평가 신뢰도를 나타내는 대표적인 계수로는 크론바흐 알파 (Cronbach's α)가 있다. 크론바흐 알파는 평가의 내적일관성 (internal consistency)을 나타낸다. 크론바흐 알파가 .80 이상이면 일반적으로 준수한 표준참조평가라고 평가한다[39].

준거참조평가의 평가 신뢰도는 통상적으로 일반화이론 (Generalizability Theory)을 기반으로 계산하나, 이수 (master)/미이수 (non-master)와 같은 이분 자료 (dichotomous data)를 분석할 때는 수식 (1)과 같이 간소화된 형태인 Brown[41]의 의존도 계수 (Dependability Φ)를 사용할 수 있다. 의존도 계수 Φ 가 .80 이상이면 일반적으로 준수한 준거참조평가라고 할 수 있다[39].

$$\Phi = \frac{\frac{nS_p^2[K-R20]}{n-1}}{\frac{nS_p^2[K-R20]}{n-1} + \frac{\bar{X}_p(1-\bar{X}_p) - S_p^2}{k-1}} \quad (1)$$

- n = 수험자 수
- k = 문항 수
- \bar{X}_p = 평균 점수
- S_p = 평균 점수의 표준편차
- K-R20 = Kuder-Richardson Formula 20 신뢰도

또한, 교육과정과 연계된 평가에서는 제곱오차손실일치도 (squared error loss agreement) $\Phi(\lambda)$ 를 사용한다. 의존도 계수와와 차이점은 기준 점수 (λ)를 지정하여 평가의 신뢰도를 확인할 수 있다는 점이다. 예를 들어, 기준 점수가 80%일 때의 평가 신뢰도와 기준 점수 90%일 때 평가 신뢰도의 변화를 확인할 수 있다. 기준점수를 평균값으로 지정하면 의존도 계수와 같은 신뢰도 값을 갖는다[39]. $\Phi(\lambda)$ 를 구하는 계산식 (2)에서 확인할 수 있듯, 문항수와 신뢰도 값은 정비례한다[41].

$$\Phi(\lambda) = 1 - \frac{1}{k-1} \left[\frac{\bar{X}_p(1-\bar{X}_p) - S_p^2}{(\bar{X}_p - \lambda)^2 + S_p^2} \right] \quad (2)$$

- k = 문항 수
- λ = 이수 기준 점수
- \bar{X}_p = 평균 점수
- S_p = 평균 점수의 표준편차

III. 웹 기반 프랑스어 지각 훈련

지각 훈련 시스템은 자바 기반의 오픈소스 실험 틀인 jsPsych (<https://www.jspsych.org>)를 이용하여 개발하였다. 분별 테스트는 주어진 두 개의 음성을 듣고 이 두 음성이 같은지 다른지를 구분하는 테스트이고, 식별 테스트는 하나의 소리를 듣고 그 소리에 해당하는 단어를 선택하는 테스트이다. 다음 그림 1은 분별 테스트(좌)와 식별 테스트(우)의 예를 각각 보여준다[42].

자음 지각 훈련 시스템은 조음위치와 조음방법을 바탕으로 최소대립쌍 475개로 이루어진 분별 (discrimination) 테스트와 최소대립쌍 233개로 이루어진 식별 (identification) 테스트로 구성하였다. 모음의 경우는 개구도와 조음위치를 바탕으로 최소대립쌍 484개로 이루어진 분별 테스트와 최소대립쌍 197개로 이루어진 식별 테스트로 구성하였다.



그림 1. 분별 테스트(좌)와 식별 테스트(우)
 Fig. 1. Discrimination Test (left) and Identification Test (right)

시스템을 구성하는 음성 파일은 프랑스어 발음 연습 교재인 Abry와 Chalaron과 Abry와 Chalaron에서 제공하고 있는 파일들 가운데 선정하였고 이러한 사용에 대해서는 개별적으로 저자의 허락을 받았다[43],[44].

지각 훈련 시스템의 테스트에는 대학에서 프랑스어를 전공으로 공부하는 학생 20명이 참여하였다. 참여자들은 모두 4년 이상 프랑스어를 학습하였으며, 실험을 위하여 프랑스어 능숙도를 따로 구분하지는 않았다. 일부 학생은 하위 테스트를 마무리하지 못하고 중도 탈락한 경우도 있었다.

IV. 지각 훈련에 대한 신뢰도 분석

앞서 제안한 <웹 기반 프랑스어 지각 훈련>은 일종의 언어 능력에 대한 평가로서 그 신뢰도를 분석하기 위해서 통상적으로 언어평가에서 활용되는 언어평가 신뢰도 평가 방법론을 채용하였다. 먼저 지각 훈련 결과의 기술 통계량을 살펴보고, 지각 훈련의 신뢰도를 분석하였다.

4-1 지각 훈련 결과의 기술 통계량

표 1은 자음 분별 테스트를 구성하는 하위 테스트 11개의 기술 통계량이다. 표 1에서 보듯, 자음 분별 테스트에 참여한 학생 수는 18~20명으로 시험 도중 일부 수험자가 탈락하였다. fax04 (sz) ~ fax07 (z3), fax09 (fp), fax11 (mn, n, wu)은 모두 1명, fax08 (bv, mn)와 fax10 (lv)은 2명이 테스트를 마무리하지 못해 분석에서 제외되었다. 문항 수는 32~52개이고, 평균 점수는 31.37~51.40으로 차이가 있으나 백분율로 환산하면 모두 90%가 넘는다. 표준편차는 0.44~3.18로 수험자가 비슷한 수준의 그룹에 속하는 것을 확인할 수 있다. 다만, fax08 (bv, mn)테스트에서는 표준편차가 3.18로 다소 큰 편이다.

표 2는 자음 식별 테스트를 구성하는 하위 테스트 6개의 기술 통계량을 보여준다. 18~19명의 수험자가 자음 인식 테스트에 참여하였다. fid04 (f3, sf)는 2명, 나머지 5개의 하위 테스트는 1명이 테스트를 마무리하지 못했다. 문항 수는 9~50개로 자음 분별 테스트보다는 적다. 평균 점수는 8.37~48.37이고, 백분율로 환산하면 자음 분별 테스트와 마찬가지로 90%가 넘는다. 표준편차는 0.68~2.63으로 자음 분별 테스트에 비해 작은 편이다. 자음 분별 테스트에서는 fax08 (bv, mn) 테스트가 표준편차가 큰 편이었으나, 자음 식별 테스트에서는 fid03 (sz, z3) 하위 테스트의 표준편차가 2.63으로 다소 큰 편이었다.

표 3은 모음 분별 테스트를 구성하는 하위 테스트 7개의 기술 통계량을 보여준다.

표 1. 자음 분별 테스트 기술 통계

Table 1. Descriptive statistics of consonant discrimination test

Subtest	n	k	Mean (%)	sd	Min	Max	Range
fax01 (pb)	20	52	51.40 (98)	0.8	50	52	3
fax02 (td,kg)	20	32	31.70 (99)	0.56	30	32	3
fax03 (fv)	20	52	51.35 (98)	1.19	48	52	5
fax04 (sz)	19	52	51.74 (99)	0.44	51	52	2
fax05 (fʒ)	19	48	47.26 (98)	0.91	45	48	4
fax06 (sf)	19	36	35.32 (98)	1.56	30	36	7
fax07 (zʒ)	19	32	31.37 (98)	1.56	27	32	6
fax08 (bv, mn)	18	40	37.44 (93)	3.18	28	40	13
fax09 (fp)	19	36	35.16 (97)	1.14	32	36	5
fax10 (lʁ)	18	52	51.33(98)	0.94	49	52	4
fax11 (mɲɲ, wu)	19	43	41.47 (96)	1.7	37	43	7

표 2. 자음 식별 테스트 기술 통계

Table 2. Descriptive statistics of consonant identification test

Subtest	n	k	Mean (%)	sd	Min	Max	Range
fid01 (pb,td,kg)	19	42	41.00 (97)	1.08	38	42	5
fid02 (fv,bv,fp)	19	50	48.37 (96)	1.9	44	50	7
fid03 (sz,zʒ)	19	42	39.74 (94)	2.63	31	42	12
fid04 (ʒ, sʒ)	18	42	41.56 (98)	0.68	40	42	3
fid05 (lʁ, wu, mn)	19	48	46.47 (96)	1.35	43	48	6
fid06 (mɲɲ)	19	9	8.37 (93)	0.81	6	9	4

표 3. 모음 분별 테스트 기술 통계

Table 3. Descriptive statistics of vowel discrimination test

Subtest	n	k	Mean (%)	sd	Min	Max	Range
fax01 (high)	17	108	106.88 (98)	1.28	104	108	5
fax02 (mid-high)	18	108	106.28 (98)	4.34	91	108	18
fax03 (mid-low)	18	53	51.67 (97)	1.8	46	53	8
fax04 (low, front, front-round)	18	28	27.78 (99)	0.42	27	28	2
fax05 (back)	18	53	51.44 (97)	2.17	46	53	8
fax06 (oral-nasal)	18	79	78.06 (98)	1.18	75	79	5
fax07 (nasal-nasal)	17	52	50.59 (97)	1.85	46	52	7

표 4. 모음 식별 테스트 기술 통계

Table 4. Descriptive statistics of vowel identification test

Subtest	n	k	Mean (%)	sd	Min	Max	Range
fid01 (high)	18	36	35.00 (97)	0.94	33	36	4
fid02 (mid-high)	18	36	35.33(98)	0.88	33	36	4
fid03 (mid-low)	18	15	14.00 (93)	1	12	15	4
fid04 (low, front, front-round)	18	14	13.89 (99)	0.31	13	14	2
fid05 (back)	18	26	19.67 (75)	2.83	16	25	10
fid06 (oral-nasal)	18	44	43.28 (98)	0.87	41	44	4
fid07 (nasal-nasal)	18	26	25.50 (98)	0.76	23	26	4

17~18명의 수험자가 모음 분별 테스트에 참여하였으며, fax01 (high)와 fax07 (nasal-nasal)는 3명, 나머지 5개의 하위 테스트는 2명이 테스트를 마무리하지 못했다. 하위 테스트의 문항 수는 52~108개로 다양하다. 평균 점수는 27.78~106.88개로, 백분율로 환산하면 모두 90%가 넘었으나, 표준편차는 0.42~4.34로 다소 작은 편이며, fax03 (mid-low) 테스트가 4.34로 상대적으로 큰 편이었다.

표 4는 모음 식별 테스트를 구성하는 하위 테스트 7개의 기술 통계량을 보여준다. 18명의 수험자가 모음 분별 테스트에 참여하였으며, 5개의 하위 테스트 전체에서 2명이 탈락하였다. 하위 테스트의 문항 수는 14~44개로 모음 분별 테스트보다는 적게 구성되어 있다. 평균 점수는 13.89~35.33개로, 백분율로 환산하면 fax05 (back)가 75%이고, 나머지는 모두 90%를 넘었다. 표준편차는 0.76~2.83으로 다소 작은 편이며, fax05 (back) 테스트가 2.83으로 상대적으로 큰 편이었다.

본 연구에 사용한 4가지 테스트, 자음 분별 테스트, 자음 식별 테스트, 모음 분별 테스트, 모음 식별 테스트의 난이도는 시험에 참여한 수험자가 대부분 90% 이상을 성취할 수 있는 수준이었다. 연구 참여자가 대학에서 프랑수를 전공으로 공부하는 학생이라는 점을 고려하면 예상되는 결과라 할 수 있다.

4-2 지각 훈련의 신뢰도 분석

본 연구에서 분석한 지각 훈련 평가의 내적일관성을 살펴 보기 위하여 먼저 크론바흐 알파값을 추출하였고, 이후 학생들의 이수/미이수를 기준으로 평가의 신뢰도를 확인하는 의존도 계수를 추출하였다.

표 5는 크론바흐 알파값과 의존도 계수를 보여준다. 본 연구에서 소개하는 <웹 기반 프랑스어 지각 훈련>은 총 4개의 테스트로 나뉘고, 31개의 하위 테스트로 이루어져 있다. 표 5에서 4개의 테스트(자음 분별, 자음 인식, 모음 분별, 모음 인식)에 해당하는 문항 전체를 하나로 묶어서 신뢰도 지표를 비교 제시하고, 표 6부터 표 9에서는 31개 하위 테스트의 신뢰도 지표를 나누어 제시하였다.

표 5에서 보듯 크론바흐 알파는 0.71~0.97로 대체로 높게 나타났고, 의존도 계수 Φ 는 0.65~0.97로 나타났다. 본 연구에 사용한 지각 훈련 평가를 상대평가 목적으로 사용하는 경우, 모음식별 테스트($\alpha=0.71$)를 제외하고는 모두 신뢰도가 높다고 볼 수 있다. 절대평가 목적으로 사용하는 경우, 자음 식별 테스트와 모음 식별 테스트의 Φ 는 0.79와 0.65로 다소 적합하지 않은 것으로 나타났다. 수행 기준을 별도로 지정하는 $\Phi(\lambda)$ 값을 살펴보면, 기준 점수가 95%인 경우에는 모음 식별 테스트의 $\Phi(\lambda)$ 가 0.61로 낮게 나타났으나, 기준 점수를 90%로 변경하면 0.91로 올라갔다. 모음 식별 테스트의 경우에는 다른 테스트에 비해 난도가 높은 편이라 기준 점수를 95%로 설정하면, 수험자의 기준 점수 당락 여부를 결정하는데 필요한 정보가 충분하지 않기 때문으로 해석할 수 있다.

기준 점수를 90%로 지정하는 경우, 4가지 테스트 모두 의존도 계수가 .90 이상으로 높게 나타났다.

표 6은 자음 분별 테스트에 속하는 하위 테스트의 신뢰도 지수를 보여준다. 상위 테스트보다 하위 테스트는 문항 수가 적어 일부 하위 테스트의 크론바흐 알파는 전반적으로 낮게 나타났다. 다만, $\Phi(\lambda)$ 의 기준 점수를 90%로 설정하면 의존도 계수가 모두 .80 이상으로 준수하게 나타났다. 기준 점수를 95%로 변경하면 fax11 (mnɲ, wɥ) 테스트의 의존도 계수가 .56까지 낮아지는 것으로 보아, 95% 기준 점수로 사용하기에는 적합하지 않음을 알 수 있다.

표 5. 4개 상위 테스트의 신뢰도 지수

Table 5. Reliability of phoneme perception tests

Subtest	Chronbach's α	Φ^*	$\Phi(\lambda=.90)^{**}$	$\Phi(\lambda=.95)^{**}$
Consonant discrimination test	.97	.97	.99	.95
Consonant identification test	.81	.79	.97	.85
Vowel discrimination test	.95	.95	.99	.97
Vowel identification test	.71	.65	.91	.61

* The phi coefficient value in high-stakes testing generally should be .80 or greater.

** In squared error loss agreement $\Phi(\lambda)$, λ represents the cut score. In this study, tests with $\lambda = .90$ and $\lambda = .95$ were compared.

표 6. 자음 분별 테스트의 신뢰도 지수

Table 6. Reliability of consonant discrimination test

Subtest	Chronbach's α	Φ	$\Phi(\lambda=.90)$	$\Phi(\lambda=.95)$
fax01 (pb)	.16	.15	.97	.87
fax02 (td,kg)	.08	.08	.96	.85
fax03 (fv)	.59	.58	.97	.88
fax04 (sz)	-.29	-.29	.98	.95
fax05 (ʃʒ)	.19	.18	.95	.79
fax06 (sj)	.76	.75	.94	.83
fax07 (zʒ)	.79	.78	.93	.83
fax08 (bv, mn)	.81	.79	.82	.79
fax09 (fp)	.40	.40	.90	.63
fax10 (lʁ)	.33	.32	.97	.85
fax11 (mnɲ, wɥ)	.54	.53	.86	.56

표 7. 자음 식별 테스트의 신뢰도 지수

Table 7. Reliability of consonant identification test

Subtest	Chronbach's α	Φ	$\Phi(\lambda=.90)$	$\Phi(\lambda=.95)$
fid01 (pb,td,kg)	.25	.23	.91	.59
fid02 (fv,bv,fp)	.64	.61	.89	.64
fid03 (sz,zʒ)	.73	.72	.81	.70
fid04 (ʃʒ, sj)	.15	.14	.97	.86
fid05 (lʁ, wɥ, mn)	.38	.33	.88	.43
fid06 (mnɲ)	.44	.34	.20	.15

표 7은 자음 식별 테스트에 속하는 하위 테스트의 신뢰도 지수를 보여준다. 자음 분별 테스트의 신뢰도 지수와 비교하여 특이한 점은 $\Phi(\lambda)$ 의 기준 점수를 90%로 설정했을 때 fid06 (mnɲ) 테스트의 의존도 계수가 .20으로 매우 낮다는 점이다. 의존도 계수는 테스트 문항의 수에 영향을 받는데, fid06 (mnɲ) 테스트는 문항이 9개로 다른 자음 인식 테스트가 40개가 넘는 것에 비하면 상당히 적은 편이다. 또한, 표 2에서 보았듯 최저 점수가 6점이고 최고 점수가 9점으로 점수의 범위가 상대적으로 크기 때문에, 기준 점수를 .90으로 했을 때 기준 점수로 수험자 그룹을 제대로 구분하기 어렵다는 것을 알 수 있다. 따라서, fid06 (mnɲ) 테스트를 사용할 때 문항 수를 늘리거나, 시험 목적을 고려하여 기준 점수를 낮게 설정하여 사용하면 테스트의 신뢰도는 높게 유지할 수 있다.

표 8은 모음 분별 테스트에 속하는 하위 테스트의 신뢰도 지수를 보여준다. 자음 분별 테스트와 마찬가지로 상위 테스트에 비해 문항 수가 적어 크론바흐 알파는 전반적으로 낮게 나타났다. 기준 점수를 90%로 설정했을 때 $\Phi(\lambda)$ 는 모두 .90 이상으로 높게 나타났다. 기준 점수를 95%로 변경하면 fax03 (mid-low), fax04 (low, front, front-round), fax07 (nasal-nasal) 테스트의 $\Phi(\lambda)$ 가 각각 .74, .75, .72로 다소 낮아졌다.

표 8. 모음 분별 테스트의 신뢰도 지수

Table 8. Reliability of vowel discrimination test

Subtest	Chronbach's α	Φ	$\Phi(\lambda=.90)$	$\Phi(\lambda=.95)$
fax01 (high)	0.36	0.36	0.98	0.94
fax02 (mid-high)	0.93	0.92	0.98	0.95
fax03 (mid-low)	0.68	0.64	0.93	0.74
fax04 (low, front, front-round)	-0.19	-0.18	0.96	0.85
fax05 (back)	0.71	0.7	0.92	0.75
fax06 (oral-nasal)	0.44	0.4	0.98	0.91
fax07 (nasal-nasal)	0.63	0.63	0.92	0.72

표 9. 모음 식별 테스트의 신뢰도 지수

Table 9. Reliability of vowel identification test

Subtest	Chronbach's α	Φ	$\Phi(\lambda=.90)$	$\Phi(\lambda=.95)$
fid01 (high)	0.15	0.12	0.87	0.36
fid02 (mid-high)	0.24	0.23	0.93	0.68
fid03 (mid-low)	0.34	0.28	0.25	0.12
fid04 (low, front, front-round)	-0.07	-0.06	0.93	0.75
fid05 (back)	0.52	0.48	0.78	0.86
fid06 (oral-nasal)	0.24	0.21	0.95	0.75
fid07 (nasal-nasal)	0.26	0.25	0.9	0.6

표 9에서는 모음 식별 테스트에 속하는 하위 테스트의 신뢰도 지수를 보여준다. $\Phi(\lambda)$ 의 기준 점수를 90%로 설정했을 때 fid03 (mid-low) 테스트의 의존도 계수가 .25으로 매우 낮게 나타났다. 표 4에서 보았듯, fid03 (mid-low) 테스트의 문항 수가 15개로 다소 적은 편이지만, fid04 (low, front, front-round) 테스트의 문항 수가 14개로 더 적으나 의존도 계수는 .93으로 높은 편으로 나타났다. fid03 (mid-low) 테스트의 기준 점수를 80%로 바꾸면 의존도 계수가 .81로 상승한다. 다만, 모음 식별 테스트의 수험자 점수가 모두 80% 이상이므로, 해당 테스트의 기준 점수 설정과 함께 테스트의 신뢰도를 평가해야 한다.

V. 논의 및 결론

본 연구는 프랑스어를 배우는 한국인 학습자를 대상으로 설계한 웹 기반의 음소지각 훈련 시스템을 소개하고 이 시스템의 신뢰도를 평가하였다. 제한한 지각 훈련은 자음 및 모음에 대한 분별 테스트와 식별 테스트로 구성되었다. 이러한 지각 훈련은 일종의 언어 능력에 대한 평가로서 그 신뢰도를 분석하기 위하여 먼저 지각 훈련 결과의 기술 통계량을 살펴보고, 준거참조평가 방법으로 지각 훈련의 신뢰도를 분석하였다.

본 연구를 위하여 자음 분별 테스트, 자음 식별 테스트, 모음 분별 테스트, 모음 식별 테스트의 4가지 테스트가 수행되었는데, 테스트 결과 난이도는 시험에 참여한 수험자가 대부분 90% 이상을 성취하였다. 지각 훈련 평가의 내적일관성을 검토하려고 먼저 크론바흐 알파 (Cronbach's alpha)값을 살펴보고, 학생들의 이수/미이수를 기준으로 평가의 신뢰도를 확인하는 의존도 계수 (dependability Φ)를 살펴보았다.

표 5에서 보았듯, 섹션별로 문항을 묶어서 평가한 테스트의 경우 대체로 신뢰도 지수가 높은 편이었으나, 모음 식별 테스트만 의존도 계수(.65)가 낮게 나타났다. 각 테스트를 발음 기준으로 나눠서 신뢰도 지수를 평가해본 결과, 문항 수가 적은 일부 하위 테스트의 의존도 계수가 낮아졌다. 의존도 계수는 문항 수에 영향을 받으므로, 문항 수가 지나치게 적은 하위 테스트는 다른 항목과 묶어서 하나의 테스트로 활용하는 편이 신뢰도가 높은 평가를 사용하는 방법이다.

아울러, 기준 점수 (λ)에 따라 의존도 계수의 변화가 있으므로, 테스트의 수준과 평가하고자 하는 수험자의 수준을 고려하여 기준 점수를 설정하는 것도 신뢰도 높은 테스트를 개발하는 데 중요한 요소가 된다. 또한, 준거참조평가의 신뢰도 기준으로 사용하는 크론바흐 알파값은 하위 테스트에서 대부분 낮게 나타났다. 이는 하위 테스트는 문항 수가 상대적으로 적기 때문이다. 이에 반해, 준거참조평가의 신뢰도 기준인 의존도 계수는 문항 수가 줄어든 하위 테스트에서도 대부분의 경우 높게 유지되었다.

이와 같은 연구는 외국어 교육에 있어서 기본이 되는 음소 습득 훈련 시스템이 개발 사례로서 다양한 언어 교육에 적용

될 수 있다. 또한, 제한한 신뢰도 평가 방법을 바탕으로 효율적이고 신뢰성 있는 테스트 세트를 구성할 수 있다. 그뿐만 아니라 좀 더 많은 학습자를 대상으로 이러한 훈련을 시행하여 축적된 데이터를 바탕으로 지능적인 학습자 맞춤형 시스템을 개발하는 데 기여할 것으로 기대된다.

감사의 글

이 논문은 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2021S1A5A2A03067523)

참고문헌

- [1] C. Puren, "De l'approche communicative à la perspective actionnelle," *Le Français dans le monde*, Vol. 347, pp. 37-40, September 2006.
- [2] J. E. Flege, Second-language speech learning: Theory, findings, and problems, in *Speech perception and linguistic experience: Theoretical and methodological issues*, Timonium, MD: York Press, ch. 8, pp. 233-277, 1995.
- [3] C. T. Best, A direct realist view of cross-language speech perception, in *Speech perception and linguistic experience: issues in cross-language research*, Baltimore: York Press, ch. 6, pp. 171- 204, 1995.
- [4] C. T. Best and M. D. Tyler, Nonnative and second-language speech perception: Commonalities and complementarities, in *Language experience in second language speech learning: In honor of James Emil Flege*, Amsterdam: John Benjamins, ch. 2, pp. 13-34, 2007.
- [5] S. Kim and H. Jung, "A study on the utilization of speech recognition technology in foreign language learning applications: Focusing on English and French speech," *Journal of Digital Contents Society*, Vol. 19, No. 4, pp. 621-630, April 2018.
<https://doi.org/10.9728/dcs.2018.19.4.621>
- [6] O. S. Bohn and J. E. Flege, "The production of new and similar vowels by adult German learners of English," *Studies in Second Language Acquisition*, Vol. 14, No. 2, pp. 131-158, June 1992.
- [7] J. E. Flege, "The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification," *Journal of Phonetics*, Vol. 15, No. 1, pp. 47-65, January 1987.
[https://doi.org/10.1016/S0095-4470\(19\)30537-6](https://doi.org/10.1016/S0095-4470(19)30537-6)
- [8] C. T. Best, Emergence of language-specific constraints in perception of non-native speech: A window on early

- phonological development, in *Developmental neurocognition: Speech and face processing in the first year of life*, pp. 289-304, 1993.
- [9] M. Sakai and C. Moorman, "Can perception training improve the production of second language phonemes? A meta-analytic review of twenty-five years of perception training research," *Applied Psycholinguistics*, Vol. 39, No. 1, pp. 1-36, October 2017.
<https://doi.org/10.1017/S0142716417000418>
- [10] R. I. Thomson, English accent coach [Computer program]. Available: <http://www.englishaccentcoach.com>
- [11] R. I. Thomson and T. M. Derwing, "The effectiveness of L2 pronunciation instruction: A narrative review," *Applied Linguistics*, Vol. 36, No. 3, pp. 326-344, December 2014.
<https://doi.org/10.1093/applin/amu076>
- [12] M. Motohashi-Siogo and D. M. Hardison, "Acquisition of L2 Japanese geminates: Training with waveform displays," *Language Learning & Technology*, Vol. 13, No. 2, 29-47, June 2009.
- [13] D. Bridges, A. Pitiot, M. R. MacAskill, and J. Peirce, "The timing mega-study: Comparing a range of experiment generators, both lab-based and online," arXiv Preprint, August 2020. <https://doi.org/10.31234/osf.io/d6nu5>
- [14] R. Ellis, "Measuring implicit and explicit knowledge of a second language: A psychometric study," *Studies in Second Language Acquisition*, Vol. 27, No. 2, pp. 141-172, June 2005. <https://doi.org/10.1017/S0272263105050096>
- [15] D. G. Jamieson and D. E. Morosan, "Training non-native speech contrasts in adults: Acquisition of the English /ð/-θ/ contrast by francophones," *Perception and Psychophysics*, Vol. 40, No. 4, pp. 205-215, June 1986.
<https://doi.org/10.3758/BF03211500>
- [16] S. E. Lively, J. S. Logan, and D. B. Pisoni, "Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories," *The Journal of the Acoustical Society of America*, Vol. 94, No. 3, pp. 1242-1255, September 1993.
<https://doi.org/10.1121/1.408177>
- [17] J. S. Logan and J. S. Pruitt, Methodological issues in training listeners to perceive nonnative phonemes, in *Speech perception and linguistic experience: Issues in cross-language research*, Timonium, MD: York Press, pp. 351-378, 1995.
- [18] J. S. Logan, S. E. Lively, and D. B. Pisoni, "Training Japanese listeners to identify English /r/ and /l/: A first report," *The Journal of the Acoustical Society of America*, Vol. 89, No. 2, pp. 874-886, Feb. 1991.
<https://doi.org/10.1121/1.1894649>
- [19] W. Strange and S. Dittmann, "Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English," *Perception and Psychophysics*, Vol. 36, No. 2, pp. 131-145, March 1984.
<https://doi.org/10.3758/BF03202673>
- [20] Y. Shinohara and P. Iverson, "High variability identification and discrimination training for Japanese speakers learning English /r/-l/," *Journal of Phonetics*, Vol. 66, pp. 242-251, January 2018.
<https://doi.org/10.1016/j.wocn.2017.11.002>
- [21] R. P. Wayland and B. Li, "Effects of two training procedures in cross-language perception of tones," *Journal of Phonetics*, Vol. 36, No. 2, pp. 250-267, April 2008. <https://doi.org/10.1016/j.wocn.2007.06.004>
- [22] K. Nishi and D. Kewley-Port, "Training Japanese listeners to perceive American English vowels: Influence of training sets," *Journal of Speech, Language, and Hearing Research*, Vol. 50, No. 6, pp. 1496-1509, December 2007.
[https://doi.org/10.1044/1092-4388\(2007\)103](https://doi.org/10.1044/1092-4388(2007)103)
- [23] K. Nishi and D. Kewley-Port, "Nonnative speech perception training using vowel subsets: Effects of vowels in sets and order of training," *Journal of Speech, Language, and Hearing Research*. Vol. 51, No. 6, pp. 1480-1493, December 2008.
[https://doi.org/10.1044/1092-4388\(2008\)07-0109](https://doi.org/10.1044/1092-4388(2008)07-0109)
- [24] R. De Leeuw, "jsPsych: A JavaScript library for creating behavioral experiments in a Web browser," *Behavior Research Methods*, Vol. 47, No. 1, pp. 1-12, March 2015.
<https://doi.org/10.3758/s13428-014-0458-y>
- [25] C. Aliaga-García and J. C. Mora, Assessing the effects of phonetic training on L2 sound perception and production, In *Recent Research in Second Language Phonetics/Phonology: Perception and Production*, Cambridge: Cambridge Scholars Publisher, ch. 1, pp. 2-31, 2009.
- [26] A. R. Bradlow, D. B. Pisoni, R. Akahane-Yamada, and Y. I. Tohkura, "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," *The Journal of the Acoustical Society of America*, Vol. 10, No. 4, pp. 299-2310, April 1997.
<https://doi.org/10.1121/1.418276>
- [27] P. Iverson and B. G. Evans, "Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers," *The Journal of the Acoustical Society of America*, Vol. 126, No. 2, pp. 866-877, July 2009. <https://doi.org/10.1121/1.3148196>
- [28] S. G. Lambacher, W. L. Martens, K. Kakehi, C. A. Marasinghe, and G. Molholt, "The effects of identification

- training on the identification and production of American English vowels by native speakers of Japanese,” *Applied Psycholinguistics*, Vol. 26, No. 2, pp. 227-247, April 2005. <https://doi.org/10.1017/S0142716405050150>
- [29] A. Lengeris, “The effectiveness of auditory phonetic training on Greek native speakers’ perception and production of Southern British English vowels,” in *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics*, Athens, Greece, pp. 133-136, August 2008.
- [30] A. Rato, “Effects of perceptual training on the identification of English vowels by native speakers of European Portuguese,” in *Proceedings of the International Symposium on the Acquisition of Second Language Speech Concordia Working Papers in Applied Linguistics*, pp. 529-546, March 2014.
- [31] J. W. Wong, “The effects of perceptual and/or productive training on the perception and production of English vowels /ɪ/ and /i:/ by Cantonese ESL learners,” in *Proceedings of the 14th Annual Conference of the International Speech Communication Association*, pp. 2113-2117, August 2013. <https://doi.org/10.21437/Interspeech.2013-501>
- [32] N. Y. Ryu and Y. Kang. “Web-based high variability phonetic training on L2 coda perception,” in *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, pp. 482-486, August 2019.
- [33] L. Pierce, Learning novel vowel contrasts: experimental methods in classroom applications, Ph.D. dissertation, University of Illinois at Urbana-Champaign, IL, May 2014.
- [34] T. Lopez-Soto and D. Kewley-Port, “Relation of perception training to production of codas in English as a second language,” in *Proceedings of Meetings on Acoustics*, Portland: Oregon, pp. 1-15, October 2009. <https://doi.org/10.1121/1.3262006>
- [35] D. Nobre-Oliveira, “Effects of perceptual training on the learning of English vowels in non-native settings,” in *Proceedings of the 5th International Symposium on the Acquisition of Second Language Speech*, Brazil: Federal University of Santa Catarina, pp. 382-389, September 2007. <https://doi.org/10.5007/2175-8026.2017v70n3p15>
- [36] N. L. Trapp and O. S. Bohn, “Training Danish listeners to identify English word-final /s/ and /z/: Generalization of training and its effect on production accuracy,” in *Proceedings of the 4th International Symposium on the Acquisition of Second Language Speech*, University of Klagenfurt, pp. 343-350, September 2000.
- [37] X. Wang and M. J. Munro, “Computer-based training for learning English vowel contrasts,” *System*, Vol. 32, No. 4, pp. 539-552, December 2004. <https://doi.org/10.1016/j.system.2004.09.011>
- [38] J. Lee, J. Jang, and L. Plonsky, “The effectiveness of second language pronunciation instruction: A meta-analysis,” *Applied Linguistics*, Vol. 26, No. 3, pp. 345-366, July 2014. <https://doi.org/10.1093/applin/amu040>
- [39] N. T. Carr, *Designing and analyzing language tests*, Oxford: Oxford University Press, 2011.
- [40] J. D. Brown and T. Hudson, *Criterion-referenced language testing*, Cambridge University Press, 2002.
- [41] J. D. Brown, “Shortcut estimators of criterion-referenced test consistency,” *Language Testing*, Vol. 7, No. 1, pp. 77-97, June 1990. <https://doi.org/10.1177/026553229000700106>
- [42] S. Kim, “Une étude sur la perception des voyelles françaises par les apprenants coréens à l’aide de paires minimales,” *Enseignement de Langue et Littérature Françaises*, Vol. 69, pp. 7-29, June 2020.
- [43] D. Abry and M. L. Chalaron, *Les 500 exercices de phonétique: niveau A1-A2*, Hachette, 2010.
- [44] D. Abry and M. L. Chalaron, *Les 500 exercices de phonétique: niveau B1-B2*, Hachette, 2011.



김선희(Sunhee Kim)

1985년: 연세대학교 불어불문학과 학사(문학사)
1986년: Université Paris VII, France (언어학석사)
1990년: EHESS, France (언어학박사)

1991년~2000년: 연세대학교 강사
2000년~2001년: L&H Korea, 책임연구원
2002년~2004년: 광운대학교 연구교수
2004년~2005년: 한국과학기술원 연구교수
2005년~2013년: 서울대학교 연구교수
2013년~2019년: 네이버(주) 수석연구원
2019년~현 재: 서울대학교 불어교육과 조교수

※ 관심분야 : 음성학(Phonetics), 음성합성(Speech Synthesis), 음성인식(Speech Recognition), 언어학(Linguistics), 자동 발음평가(Automatic Pronunciation Assessment)



류나영(Na-Young Ryu)

2001년: 광운대학교 경영학과 학사
2010년: 한국외국어대학교 영어학과 (영어학석사)
2019년: University of Toronto, Canada (언어학박사)

2019년~현 재: Pennsylvania State University

※ 관심분야 : 음운론(Phonology), 음성학(Phonetics), 제2언어습득(Second Language Acquisition), 컴퓨터 지원 언어 학습(Computer-assisted Language Learning)



원용국(Yongkook Won)

2001년: 한국외국어대학교 영어과 (문학사)
2009년: 국제영어대학원대학교 영어교재개발학과 (영어학석사)
2019년: Iowa State University, USA (응용언어학박사)

2019년~현 재: 서울대학교 교육융합연구원 객원연구원

2021년~현 재: 국제영어대학원대학교 강사

※ 관심분야 : 언어평가(Language Assessment), 자동말하기/글쓰기평가(Automated Speaking/Writing Assessment), 맞춤형학습(Adaptive Learning)