



# Perception of Japanese word-initial stops by native listeners\*

Hi-Gyung Byun\*\*

*Akita International University, Akita, Japan*

## Abstract

It is known that the voicing contrast for Japanese word-initial stops is primarily realized as differences in the voice onset time (VOT). However, recent studies have reported that voiced stops are more often produced with a positive VOT than with a negative VOT among the younger generation nationwide. It is also known that post-stop F0 is associated with the stop contrast, but the degree of F0 use differs from region to region. This study explores whether the difference in post-stop F0 functions as a perceptual cue to the stop contrast along with VOT. Fifty-five college students who are native listeners from four different regions participated in two or three perception tests. The results show that VOT is a primary cue to the voiced-voiceless distinction of word-initial stops, but that the effect of post-stop F0 on the stop contrast is marginal. The post-stop F0 is involved in perception only when VOT is ambiguous, such that a sound with high F0 is more often perceived as a voiceless stop, but not vice versa. The results of this study indicate that the acoustic parameters associated with the stop contrast are not the same in production and perception, and suggest that other factors such as context, which is not an acoustic characteristic, may also be involved in the stop contrast.

**Keywords:** Japanese, word-initial stop, perceptual cue, VOT, post-stop F0

## 1. 서론

일본어의 어두 폐쇄음은 파열 전에 성대 진동을 수반하는 유성음과 성대 진동을 수반하지 않는 무성음의 2항 대립을 이루는 것으로 알려져 있다. VOT(voice onset time)으로 기술하면 유성음은 음 값, 무성음은 양 값을 갖는다. 다만 무성음의 VOT는 프랑스어나 스페인어처럼 성대 진동의 유무로 유성음(voiced/pre-voicing), 무성음(voiceless/unaspirated)을 구별하는 유성성 언어(true voice language)의 무성음보다는 길고, 영어나 독일어처

럼 기식의 유무로 무성음(voiceless/aspirated), 유성음(voiced/unaspirated)을 구별하는 기식성 언어(aspirating language)의 무성음보다는 짧아서 Lisker & Abramson(1964)의 양 값의 VOT를 갖는 두 카테고리(short lag: 0-25 ms, long lag: 60-100 ms)의 중간에 위치한다(Kong et al., 2012; Riney et al., 2007; Takada, 2011).

일본어 유성음의 VOT에 대해서는 선행연구 사이에서 의견이 엇갈린다. 어두에서 충분히 긴 음 값으로 실현된다는 보고가 있는가 하면(Itoh et al., 1979; Kobayashi, 1981; Shimizu, 1996; 1999) 유성음의 VOT는 무성음의 VOT와 매우 가까워서 VOT가

\* This work was supported by JSPS KAKENHI Grant Numbers JP19K00634, JP20K00642.

\*\* byun@aiu.ac.jp, Corresponding author

Received 30 July 2021; Revised 12 September 2021; Accepted 13 September 2021

© Copyright 2021 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

유성/무성을 구별하는 유효한 음향 변수가 될 수 없다고 주장하는 연구도 있다(Homma, 1980). 후자와 유사한 견해는 최근의 논문에서도 찾아볼 수 있는데 Kong et al.(2012)은 유성음과 무성음의 VOT가 대부분 중복된 데이터를 제시하고 VOT만으로는 두 개의 음운 카테고리를 구별하기 어렵다고 주장하였다.

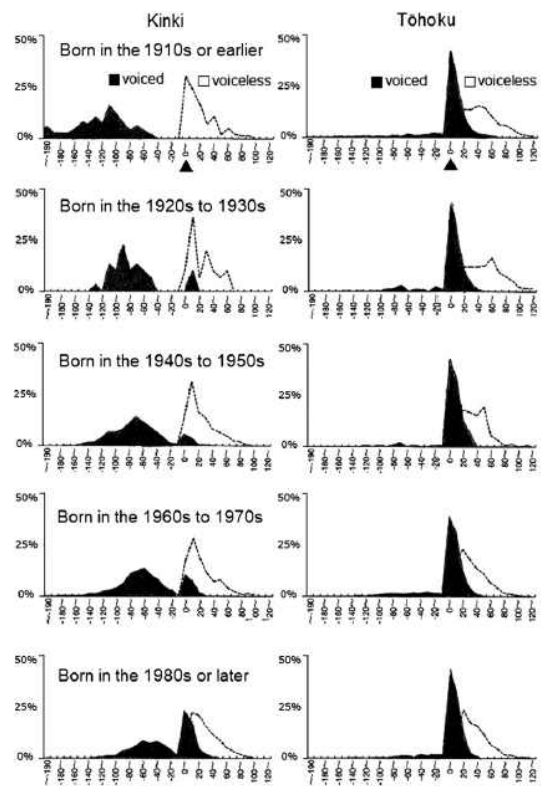
선행연구에 보이는 이 같은 불일치는 Takada(2004; 2006; 2011)에 의해 화자의 방언과 나이에 기인하는 것으로 확인되었다. Takada(2011)은 전국을 대상으로 수집된 대규모 음성 코퍼스를 이용하여 1895-2001년에 출생한 908명의 VOT를 분석하였다. 이에 따르면 전통적인 발음에서는 도호쿠지방과 간토지방 사이를 경계로 도호쿠지방과 그 북쪽에 위치하는 홋카이도는 유성음이 양 값의 VOT, 도쿄를 중심으로 하는 간토지방과 그 이서 지역은 유성음이 음 값의 VOT를 갖는다. 내용을 조금 더 자세히 보기 위해 그림 1에 Takada(2011)의 결과를 제시하였다(원본을 일부 수정. 영문 지역명 추가, 첫 번째 그림 가로축의 0 ms에 ▲ 표시 추가, 유성음을 검은색으로 표시). 가로축은 VOT의 실측치(20 ms 간격), 세로축은 전체 데이터 중에서 해당 VOT로 나타난 비율이다. 각 그림은 화자의 생년을 20년 간격으로 나누는 것이다. 왼쪽의 Kinki는 교토, 오사카를 포함하는 지역으로 간토지방과 그 이서지역의 대표지역이며 오른쪽의 Tohoku는 도호쿠지방과 홋카이도의 대표지역이다.

그림 1의 긴키지방을 보면 1910년 이전 출생자의 발화는 유성음은 파열 전에 성대 진동이 있고 무성음은 약한 기식이 있는 전형적인 유성성 언어(true voice language)로 볼 수 있다. 그러나 1920년 이후 출생자부터 유성음의 VOT가 양 값으로 나타나기 시작하여 1980년대 이후에 출생한 화자에서는 유성음의 VOT의 반 정도가 양 값으로 실현되고 있다. 이에 반해 무성음의 VOT는 생년에 따른 변화가 거의 없어 1980년대 이후 출생자는 결과적으로 유성음과 무성음의 VOT가 중복하게 된다.

오른쪽의 도호쿠지방은 화자의 생년과 상관없이 유성음의 VOT가 거의 양 값으로 나타나며 무성음의 VOT와 일부가 중복되어 있다. 본래 양 값인 무성음의 VOT는 1910년대 이전에는 VOT의 피크가 40 ms이었던 것이 생년이 내려갈수록 VOT가 조금씩 줄어 1960년대 이후가 되면 피크가 20 ms 전후까지 줄면서 유성음과의 중복의 정도가 더욱 커진다.

그림 1에서 보는 것처럼 1980년대 이후의 젊은 세대는 어느 지역이나 유성음과 무성음의 VOT가 중복하여 구별이 어려운 상태이다. 그러나 이러한 상황에서도 필자가 아는 바로 일상생활에서 발음상의 문제나 청각상의 혼란이 있다는 보고는 아직까지 없다. 이것은 VOT 이외의 다른 변수로 어두 폐쇄음이 구별되고 있는 것을 시사한다.

VOT 이외에 일본어 어두 폐쇄음의 구별에 관여하는 음향특성이 무엇인가에 대해 지금까지 후속모음의 F0, 모음의 음질(H1-H2) 등이 검토되어 왔다(Byun, 2021; Gao & Arai, 2019; Kang & Kong, 2014; Takada et al., 2016).



VOT, voice onset time.

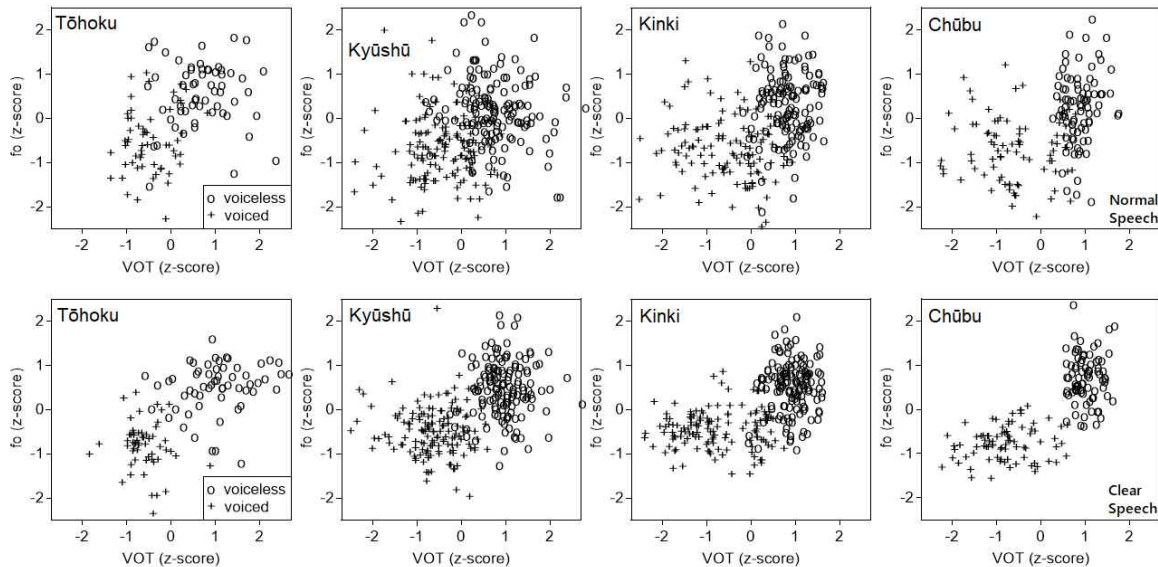
Takada (2011: Fig. 3-24), 일부 수정, Kurosio 출판 동의

그림 1. 일본어 어두 폐쇄음의 VOT 분포

Figure 1. Distribution of VOT in Japanese word-initial stops

Kong et al.(2012)는 도쿄방언 모어화자 20명의 VOT, F0, H1-H2값을 측정하여 이항로지스틱 회귀분석의 결과를 이용하여 예측의 정확도(유성/무성을 바르게 예측할 확률)를 보고하였다. 이에 따르면 남녀의 결과가 약간 다른데 먼저 여성의 발화는 VOT만으로 유성/무성을 81%의 정확도로 구별할 수 있으며 이에 F0가 더해지면 정확도가 90%로 올라간다. 여기에 다시 H1-H2가 더해지면 92%까지 개선된다. VOT와 H1-H2만으로는 예측 정확도가 84%에 그쳐서 H1-H2 효과보다는 F0 효과가 큰 것으로 나타났다. 남성의 발화는 VOT만으로 89%의 정확도로 유성/무성을 구별할 수 있으며 여기에 F0를 더해도 정확도에 변화는 없고 H1-H2를 더했을 때 91%, VOT에 F0와 H1-H2를 모두 더했을 때 92%까지 개선되는 것으로 나왔다. 정리하면, 음향 변수 VOT, F0, H1-H2를 모두 사용해서 유성/무성을 바르게 구별할 수 있는 예측 정확도는 남녀 모두 92%로 같고 일차적 음향 변수가 VOT인 것도 남녀 모두가 일치하나 F0와 H1-H2의 사용에 있어서는 남녀 차이가 있어 남성보다 여성에게서 F0 효과가 크게 나타났다. H1-H2 효과는 남녀 모두에게 인정되기는 하나 다른 변수에 비하면 그 정도는 크지 않았다.

Byun(2021)은 같은 통계 수법을 이용하여 4개 지역의 대학생(1990년대 출생) 82명을 대상으로 VOT와 F0를 분석하였다. 논문에서는 보통의 약화된 발화(normal speech)보다 음향특성이 분명히 나타나는 강화된 발화(clear speech)의 환경을 설정하여



VOT, voice onset time.

Byun (2021: Fig. 10), Phonetic Society of Japan 출판 동의

그림 2. 일본어 어두 폐쇄음의 VOT-F0 공간

Figure 2. Acoustic space of word-initial stops in Japanese (top: Normal Speech, bottom: Clear Speech)

VOT와 F0 중에서 어느 것이 강화(enhancement)되는가를 비교하였는데 결과가 방언에 따라서 달리 나왔다.

그림 2에 Byun(2021)의 결과를 제시하였다. 유성음의 VOT가 양 값으로 나타나는 비율(유성 폐쇄음의 무성화율)이 큰 순서로 왼쪽에서부터 도호쿠, 규슈, 긴키, 주부지방의 결과이다. 어두 유성 폐쇄음의 무성화율은 상단의 약화된 발화(normal Speech)에서 도호쿠 83%, 규슈 73%, 긴키 45%, 주부 26%로 보고하고 있다. 유성음의 VOT가 본래 양 값으로 나타나는 도호쿠지방을 별개로 하면 무성화율은 규슈지방에서 가장 높고 주부지방에서 가장 낮다. 달리 말하면 어두 폐쇄음의 VOT 변화에 있어서 가장 혁신적인 것은 규슈지방이고 가장 보수적인 것은 주부지방이라 할 수 있다.

Byun(2021)에 사용된 약화된 발화(normal Speech)는 1음절 무의미어의 단독발화, 강화된 발화(clear Speech)는 같은 1음절을 회화체로 묻고 답하는 형식이었다. 일반적인 의사소통 환경을 고려한다면 회화체의 강화된 발화가 보통(normal)이고 단독발화의 약화된 발화가 실험실 환경이 될 것이다. 그림 2에서 보듯이 어느 지방이나 약화된 발화에 비해 강화된 발화에서 유성음과 무성음의 구별이 분명해지는데 강화된 발화에서 강화된 음향 변수는 도호쿠지방과 긴키지방은 F0, 규슈지방과 주부지방은 VOT였다. 어느 지방이나 VOT 단독일 때보다 F0가 더해질 때 예측 정확도가 높아지나 규슈지방과 주부지방은 약화된 발화에서 F0가 더해질 때보다도 강화된 발화에서 VOT 단독일 때가 정확도가 더 높아서 발화 효과(Normal vs Clear)가 F0 효과보다도 크게 나왔다. 유성/무성을 정확히 예측할 수 있는 확률은 리스트 읽기인 약화된 발화에서 VOT만으로는 81%~91%이나 회화체인 강화된 발화에서는 도호쿠, 규슈, 긴키지방은 93%~95%까지 올라가며 여기에 F0가 더해지면 세 지역의 예측 정확

도는 97%까지 올라가 실질적으로 VOT와 F0만으로 유성/무성을 거의 구별할 수 있게 된다(VOT 효과가 큰 주부지방은 회화체에서 VOT만으로 99% 구별이 가능).

이처럼 산출에서는 VOT 외에 후속모음의 F0가 폐쇄음의 구별에 관여하고 있는 것이 분명한 듯하나 지각에서도 F0가 관여하고 있는지에 대해서는 충분히 검토되어 있지 않다. 지각 변수가 무엇인지를 논할 때 산출에서 관찰되는 과일 전의 성대 진동(buzz bar)은 언제나 연구의 중심이 되어왔다. 그러나 성대 진동의 유무가 유성음의 지각에 관여하기는 하되 직접적인 요인이 아니라는 것은 예나 지금이나 일관된 견해이다(Matsui, 2014; Otomo et al., 1957; Sato, 1958; Takada, 2004).

산출에 사용되는 음향특성이 지각에도 사용될 것이라는 예상은 통상적인 추론이다. 다만 산출에서 관찰되는 음향특성이 지각에서 관찰되지 않는 경우가 있는데 한국어 서울말의 경우도 그 중 하나이다. 서울말의 어두 경음은 산출에서는 짧은 VOT와 높은 F0로 나타나지만 모어화자(=청자)는 지각에서 F0를 거의 이용하지 않고 주로 VOT만으로 경음을 지각하는 것이 알려져 있다(Kim, 2004; Kong, 2012; Lee et al., 2013).

반대로 산출에서 사용되지 않는 음향특성이 지각에서 사용되는 경우도 있다. 영어의 어두 폐쇄음은 산출에서는 VOT만으로 충분히 구별이 되며(Abramson & Lisker, 1970; 1973; Abramson & Whalen, 2017; Kong et al., 2012) 지각에서도 VOT 차이가 분명할 때는 다른 음향특성을 필요로 하지 않으나 VOT 차이가 애매할 때는 후속모음의 F0 차이가 지각특성으로 쓰이는 것이 알려져 있다(Abramson & Lisker, 1985; Winn et al., 2013; cf. Whalen et al., 1993).

본고의 목적은 일본어의 어두 폐쇄음 지각에 VOT 외에 F0가 관여하는지를 밝히는 것이다. 산출에서 정도의 차이를 감안하

더라도 어느 지역이나 F0가 유성/무성의 구별에 관여하고 있으므로 지각에서도 그 정도에 따라 F0의 관여가 있을 것이라는 예상하에 (1) 자연음과 (2) 합성음을 이용한 총 세 가지의 지각 실험을 실시하였다. 결과는 예상과는 달리 (1) 무성화한 유성음(VOT가 양 값)의 지각은 모어 청자라도 판별이 어려우며 (2) 지역과 상관없이 폐쇄음 지각에 F0의 관여는 극히 한정적이었다.

이하에서는 2.에서 실험 방법에 대해 설명하고 3.에서 결과를 보고한다. 4.에서 결과를 바탕으로 지각에 관여하는 음향 변수 및 그 외의 요인에 대해 논의하고 5.에서 결론으로 마무리한다.

## 2. 지각 실험

실험 1은 자연음, 실험 2, 3은 합성음을 사용하였다. 실험 1은 VOT가 양 값인 유성음을 모어청자가 바르게 인지하고 있는지를 확인하기 위한 것이고 실험 2, 3은 VOT와 F0만을 변수로 하여 각각의 효과를 확인하기 위한 것이다. 실험 2는 VOT가 양 값만, 실험 3은 VOT가 양 값과 음 값을 모두 포함하는 자극음을 사용하였다.

실험 1, 2는 청자 그룹이 동일하여 도호쿠, 간토, 간키, 주부지방의 대학생, 실험 3은 주부지방의 대학생만 참가하였다. 참고로 실험 1, 2는 서울방언 모어청자 34명에게도 실시하였다.

### 2.1. 실험 1 자연음 /pa/ vs /ba/

#### 2.1.1. 자극음

그림 2의 Byun(2021)에서 쓰인 무의미어의 단독 발화(normal speech) 328개(82명이 /pa/ /ba/를 2번씩 발화) 중에서 VOT가 0-25 ms인 168개(그림 3의 □표시)와 더미(dummy) 12개를 더한 180개를 자극음으로 사용하였다. 자극음 168개는 /pa/가 77개로 VOT는 6-25 ms, /ba/가 91개로 VOT는 4-24 ms이다. 그림 3은 남녀의 발화가 섞여 있는데 실험에서도 그대로 사용하였다. 일어날 수 있는 혼란을 막기 위해 실험 시에 제시한 지시문에 남녀 목소리가 섞여 나오는 것을 명기하고 연습에서도 남녀의 예를 섞어서 제시하였다.

그림 4는 그림 3에서 선정한 자극음을 VOT와 F0로 나누어 나타낸 것이다. VOT의 중앙치는 /ba/가 10 ms, /pa/가 16 ms로 무성음의 VOT가 약간 긴 음으로 구성되어 있다. F0는 남녀의 음성이 섞여 있어 분포가 둘로 나뉘어져 있다. /pa/가 /ba/보다 F0 값이 약간 높아서 남녀 모두 중앙치로 12 Hz 차이가 있다. 유성음과 무성음 사이의 이러한 차이가 결과에 영향을 미칠 수도 있겠으나 결과에서는 위의 차이에 기인한다고 여겨지는 특징은 관찰되지 않았다. 그림 4에는 없으나 더미의 VOT는 -57- -51 ms로 결과에서는 더미에 대해서도 언급하겠다.

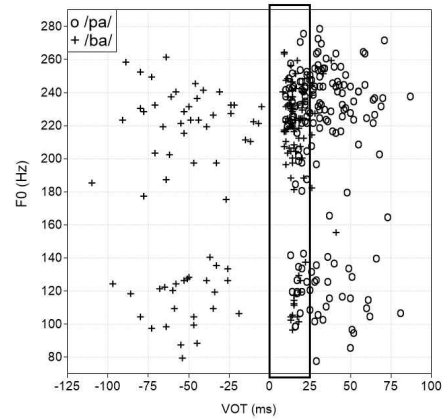
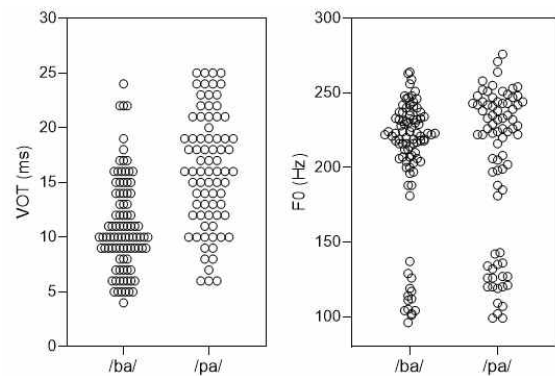


그림 3. 실험 1의 자극음  
Figure 3. Stimuli for Exp. 1



VOT, voice onset time.

그림 4. 자극음의 VOT와 F0 분포  
Figure 4. VOT and F0 values of stimuli for Exp. 1

#### 2.1.2. 참가자

청자는 1997-2001년 출생의 대학생 55명이다. 지역별로는 도호쿠(홋카이도를 포함) 12명, 간토 20명, 간키 10명, 주부 13명이다. 위의 Byun(2021)의 화자와는 다른 그룹으로 지역마다 남녀의 비율이 고르지 않으므로 남녀를 나누지 않고 같이 취급하였다. 서울방언 참가자 34명은 1993-2001년 출생의 대학생이다.

### 2.2. 실험 2 합성음 /to/ vs /do/

#### 2.2.1. 자극음

자극음은 VOT가 양 값을 가지는 240개이다(자극음 108개와 더미 12개를 2회 반복). 원음은 도호쿠지방 아키타 출신의 여성(1998년 출생)이 발화한 /to/(34 ms, 240 Hz)이다. Byun(2021)의 도호쿠 화자 18명의 발화를 참고로 VOT와 F0의 범위를 정한 후 VOT는 8-72 ms를 8 ms 간격으로 9단계, F0는 170-280 Hz를 10 Hz 간격으로 12단계 설정하였다. Praat(Boersma & Weenink, 2019, version 6.0.22)의 Manipulate기능을 이용하여 먼저 VOT를 9단계로 조작한 후에 각각의 F0를 12단계로 조작하였다. 마지막으로 단조로움에서 오는 부자연스러움을 피하기 위해 조작한 음성

의 F0를 일률적으로 5% 낮추었다(onset/off set=1.05). 하강폭 5%는 다음과 같이 정하였다.

실험 1에서 사용한 자연음 180개 중 말미가 상승조인 예를 제외한 137개의 하강폭(중간지점과 말미의 F0 차이)을 측정하여 비율로 평균치 1.06, 중앙치 1.04를 얻었다. /p/, /b/ 모두 하강폭이 매우 크게 나타난 발화가 있어서(약 15%) 평균치를 그대로 사용하지 않고 평균치와 중앙치의 산술 평균인 1.05를 사용하였다(하강폭 5%). 자극음 전체에 피치 변화를 줌으로써 합성음 특유의 로봇과 같은 인상을 조금이나마 완화할 수가 있다. 다만 이러한 조작이 분석 변수에 영향을 주어서는 안 되기 때문에 하강폭은 음절초의 F0를 피해 중간지점을 택하고, 모든 음에 일률적으로 적용하였다.

실험 2의 더미는 실험 1의 자연음이므로 결과에서는 언급하지 않았다.

### 2.2.2. 참가자

실험 1의 청자와 같은 일본의 4개 지역 대학생 55명이다. 서울방언 청자는 실험 1과 같은 34명이지만 한 명은 다른 사람이다.

### 2.3. 실험 3 합성음 /ta/ vs /da/

#### 2.3.1. 자극음

자극음은 VOT가 양 값과 음 값을 가지는 264개이다(자극음 132개를 2회 반복, 더미 없음). 실험 2에서 양 값의 VOT를 갖는 자극음만으로 유성/무성의 구별이 안 되었기 때문에 음 값의 VOT를 갖는 자극음을 추가하였다. 원음은 긴키지방의 효고 출신 여성(1995년 출생)이 발화한 /ta/(30 ms, 227 Hz)와 /da/(-36 ms, 253 Hz)이다. Byun(2021)에서 유성 폐쇄음의 VOT가 음 값이 주류인 긴키지방과 주부지방의 화자 38명의 발화를 참고로 VOT와 F0의 범위를 정한 후 VOT는 -56-24 ms를 8 ms 간격으로 11단계(0 ms는 +값과 -값을 포함), F0는 170-270 Hz를 10 Hz 간격으로 11단계를 설정하였다. 실험 2와 마찬가지로 Praat의 Manipulate기능을 이용하여 먼저 VOT를 11단계로 조작한 후에 F0를 11단계로 조작하였다. 그 후에 단조로움을 피하기 위해 실험 2와 마찬가지로 자극음의 F0를 일률적으로 5% 낮추었다(onset/off set=1.05).

지각 실험은 같은 날에 연속해서 실시되었는데 단음절의 자극음은 특히 합성음의 경우 컨트롤하는 음향 변수 이외에 음의 변화가 없기 때문에 참가자가 계속해서 단조로운 자극음에 노출되면 상당한 피로감을 느낀다. 피로감에서 오는 집중력 저하와 그에 따른 데이터의 신뢰도 저하를 최소화하기 위해 전체의 피치를 일률적으로 낮춘 것 외에도 실험에 쓰이는 음의 종류를 달리하여 변화를 주었다. 다만 이미 언급한 것처럼 이러한 변화가 실험 결과에 영향을 미쳐서는 안되므로 자음은 VOT 차이가 거의 없는 양순음과 치경음을 사용하고 모음은 F0 차이가 상대적으로 큰 고모음을 피해 /a/와 /o/를 사용하였다. 모음에 대해서는 Byun(2021)에서 후속모음이 /i/일 때는 무성음의 F0가 높아지는 경향이 있으나 모음의 종류가 유성/무성의 카테고리 구별

과는 무관한 것이 이미 확인된 바 있다. 따라서 본 연구에서도 자음과 모음의 차이가 유성/무성의 인지에 직접적인 영향을 미치지 않을 것이라는 전제하에 실험을 진행하였다.

### 2.3.2. 참가자

실험 1, 2에 참가한 주부지방 출신의 13명이다.

### 2.4. 실험 수순

도호쿠, 간토, 긴키지방의 참가자는 실험 1, 2를 2019년 6-7월, 주부지방의 참가자는 실험 1, 2, 3을 2019년 10월에 마쳤다. 본실험 전에 본실험과 동일한 형식으로 8개의 자극음을 사용한 연습을 실시하였다. 실험은 연습 1 → 실험 1 → 연습 2 → 실험 2 → (연습 없음) → 실험 3의 순서로 진행하였다.

실험장소는 대학교의 연구실로 참가자는 실험 전에 실험에 관한 구두설명을 들었다. 그 후에 노이즈 캔슬 기능이 있는 헤드폰(Sony WH-1000XM3, Tokyo, Japan)을 착용하고 헤드폰에서 들려오는 음성과 가장 가깝다고 생각되는 음성을 컴퓨터 화면에 표시되는 2개의 선택지에서 클릭하여 선택하였다. 다시 듣고 싶을 때는 “다시 한번”을 클릭하여 다시 들을 수 있도록 하였고 잘못 선택하여 수정하고 싶을 때는 “하나 앞으로”를 클릭하여 앞으로 돌아가서 다시 들을 수 있게 설정하였다.

실험 1, 2는 30문제마다 실험 3은 40문제마다 화면에 휴식하라는 지시문이 나타나 실시 유무는 참가자의 재량에 맡겼다. 다만 실험과 실험 사이는 헤드폰을 벗고 휴식을 취하도록 하였다.

그림 5는 위에서부터 실험 1(pa-ba), 실험 2(to-do), 실험 3(ta-da)의 화면이다. 일본어로 듣는 것이므로 선택지는 가나표기를 하였다. 실험 1, 2에 참가한 서울방언 청자의 선택지는 영문 표기를 하였고 지시문에는 한글로 “음성은 외국어입니다”라고 명기하였다.

모든 참가자는 실험 후에 소정의 상품권을 사례로 받았고, 실험 결과의 공개 동의서에 자필 서명하였다.



그림 5. 실험 1(상단), 실험 2(중앙), 실험 3(하단)의 화면  
Figure 5. Experiment screens for the Exp. 1 (top), Exp. 2 (center) and Exp. 3 (bottom)

### 3. 결과

#### 3.1. 실험 1 자연음 /pa/ vs /ba/

표 1과 그림 6에 실험 1의 결과를 제시하였다. 표 1의 평균 정답률은 먼저 무성음 /pa/가 94%였다. 지역별로는 93%~95%로 지역 간 차이가 거의 없고 그림 6에서 보는 것처럼 3명을 제외하고는 어느 지역이나 대개가 90% 이상 맞게 답하였다. 이에 반해 유성음 /ba/의 평균 정답률은 66%로 상대적으로 매우 낮고 지역별로 봐도 63%~69%로 저조하다. 선택지가 두 개인 것을 고려하면 유성음의 판별이 거의 안 되고 있다고 할 수 있다. 정답률이 50%를 넘지 못한 이들도 여럿 포함되어있는 것을 그림 6에서 확인할 수 있다. 유성음 /ba/의 VOT는 모두 양 값으로 산출의 경우와는 달리 유효한 VOT 정보가 없으면 유성음의 지각이 어려운 듯하다.

표 1. 자연음 /pa/ /ba/의 지역별 정답률  
Table 1. Correct response rates of /pa/ /ba/ by region

	주부 Chubu (%)	간토 Kanto (%)	긴키 Kinki (%)	도호쿠 Tohoku (%)	평균 mean (%)
/pa/	95	93	93	93	94
/ba/	63	67	69	64	66

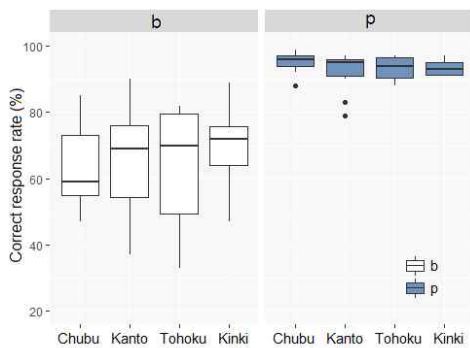


그림 6. 자연음 /pa/ /ba/의 지역별 정답률

Figure 6. Correct response rates of natural sounds /pa/ and /ba/ by the native Japanese listeners of four regions

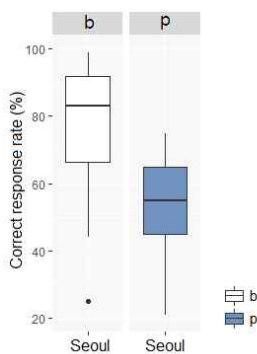


그림 7. 자연음 /pa/ /ba/의 정답률 (서울방언 참가자)

Figure 7. Correct response rates of natural sounds /pa/ and /ba/ by the native listeners of Seoul Korean

자음 종류(유성음, 무성음)와 지역(4개 지역)을 독립변수, 정답률을 종속변수로 하는 이원배치분산분석을 실행한 결과는 상호작용과 지역의 주효과는 유의하지 않았으나 [ $F(3, 50)=0.4818$ ,  $p=0.6964$ ,  $F(3,5)=0.3847$ ,  $p=0.7645$ ], 자음 종류는 무성음의 정답률이 유성음의 정답률보다 유의하게 높았다 [ $F(1, 50)=129.5$ ,  $p<0.0001$ ]. 산출에서 거의 VOT만으로 유성/무성을 구별하는 주부지방은 다른 지역보다 /pa/의 정답률이 높고 /ba/의 정답률이 낮아 지각에서도 다른 지역보다 VOT의 의존도가 높은 것으로 보이나 통계적으로는 유의하지 않았다.

실험 1에 쓰인 더미 /ba/의 VOT는 -57--51 ms로 정답률은 어느 지역이나 100%였다. 이 결과만 보면 파열 전의 성대 진동이 유성음의 지각 변수로 보이지만 문제는 서론에서도 언급했듯이 성대 진동이 없어도 유성으로 지각되므로 유성음 지각의 직접적인 변수로 보기 어렵다는 점이다.

그림 7은 일본어를 서울방언 참가자가 청취한 결과이다. 평균 정답률은 무성음 /pa/가 50%, 유성음 /ba/가 77%로 일본인 청자와는 반대로 유성음의 정답률이 상대적으로 높았다. 무성음의 정답률 50%는 무성음의 구별이 전혀 안 되고 있는 것을 나타낸다. 그림 3에 있는 것처럼 자극음의 F0 값은 유성/무성의 차이가 거의 없어 지각 변수로써 도움이 되지 않기 때문에 전적으로 VOT에 의존해야 하는데 무성음 /pa/의 VOT가 무성음으로 판정하기에는 충분히 길지 않아 판단이 어려웠던 것으로 추정된다.

#### 3.2. 실험 2 합성음 /to/ vs /do/

그림 8에 실험 2의 결과를 제시하였다. 지역 방언과 상관없이 어느 지역의 청자나 VOT가 양 값인 자극음을 모두 무성음으로 지각하였다. 8 ms의 극히 짧은 VOT일때 유성음 지각이 약간 우세하게 나타났는데 같은 8 ms이더라도 유성음 지각이 우세한 것은 F0가 낮을 때만으로 높은 F0에서는 예외 없이 무성음으로 지각하였다. 이것은 VOT가 극히 짧아 지각 변수로써 제대로 기능하지 못할 경우 F0가 유성/무성의 구별에 쓰일 수 있다는 것을 시사한다.

서론에서 언급한 것처럼 산출에서는 일본어 유성 폐쇄음의 VOT가 양 값으로 나타나는 현상을 정도의 차이는 있어도 어느 지역에서나 쉽게 찾아볼 수 있다. 그러나 지각에서는 양 값의 VOT를 갖는 자극음을 유성음이 아닌 무성음으로 판단하고 있어 VOT 사용에 있어서 산출과 지각의 결과가 일치하지 않는 것을 보여준다.

그림 9는 서울방언 참가자의 결과이다. 일본인 청자와는 완전히 다른 양상으로 긴 VOT와 높은 F0는 무성음, 짧은 VOT와 낮은 F0는 유성음으로 지각하고 있는 것을 알 수 있다. 일본인 청자가 전적으로 VOT에 의존하여 지각하는 것과는 달리 서울방언 참가자는 VOT와 F0를 거의 균등하게 사용하고 있다고 할 수 있다.

실험 2는 서울방언 참가자에게 먼저 실시하였는데 그림 9에서 보는 것처럼 VOT가 64 ms에서도 유성음과 무성음의 경계치를 확인할 수 없었기 때문에 그림 8에서는 VOT를 한 단계 늘린 72 ms를 추가하였지만 일본어에서는 무의미한 일이었다.

서울방언 참가자의 결과는 실험 2의 자극음에 문제가 없는 것을 증명해 준다. 그림 8만으로는 자극음 자체에 문제가 있는 것이 아닌가 의심을 들게 하는 의외의 결과이기 때문이다.

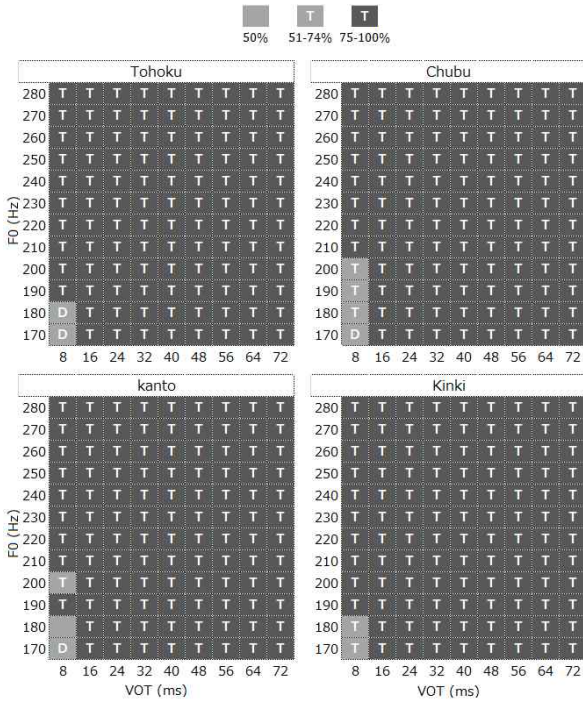


그림 8. 합성음 /to/ /do/의 인식률(일본어 4개 방언)  
Figure 8. Identification rates of synthesized sounds /to/ and /to/ by the native Japanese listeners of four regions

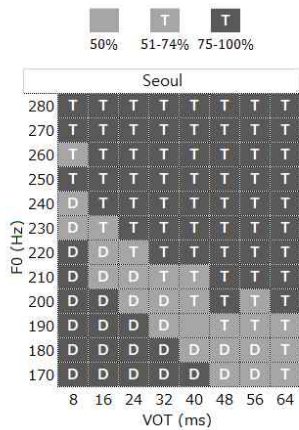


그림 9. 합성음 /to/ /do/의 인식률(서울방언)  
Figure 9. Identification rates of synthesized sounds /to/ and /to/ by the native listeners of Seoul Korean

### 3.3. 실험 3 합성음 /ta/ vs /da/

실험 3의 결과를 그림 10과 그림 11에 제시하였다. 실험 2에서는 음 값의 VOT를 취급하지 않았기 때문에 실험 3에서는 음 값과 양 값의 VOT를 모두 포함하였다. 그림 10은 실험 2와 마찬가지로 인식률을 50% 이하와 75% 이상으로 나누는 것이고 그림 11은 F0의 관여를 더 자세하게 보기 위해 인식률을 10포인트 단

위로 상세히 나눈 것이다.

그림 10을 먼저 보면 음 값의 VOT는 유성음으로, 양 값의 VOT는 무성음으로 지각하고 있는 것이 명백하다. 다만 선행연구에서도 언급하였듯이 음 값의 VOT가 유성음 지각에 필수적이지 않은 것을 고려하면 ‘양 값의 VOT가 아닌 음’을 유성음으로 지각하고 있을 가능성도 있다. 다시 말해 음 값의 VOT를 유성음, 양 값의 VOT를 무성음으로 지각하는 것이 아니고 양 값의 VOT를 무성음, 양 값이 아닌 VOT를 유성음으로 지각하고 있을 가능성이 있다. VOT가 0 ms인 경우(+0 ms는 원음의 VOT가 양 값, -0 ms는 원음의 VOT가 음 값인 음을 0 ms로 조작한 것)에는 F0가 낮을 때 무성음 지각이 줄고 F0가 높을 때 유성음 지각이 약간 주는 것으로 보아 F0의 관여가 의심되나 이것만으로는 F0가 어떻게 관여하고 있는지 알 수 없으므로 그림 11에서 자세하게 보도록 하겠다.

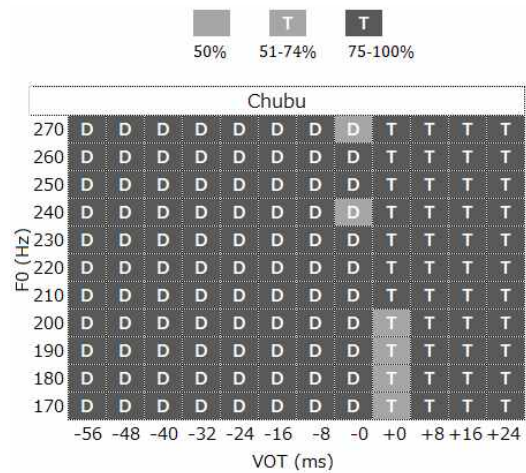


그림 10. 합성음 /ta/ /da/의 인식률(주부지방)  
Figure 10. Identification rates of synthesized sounds /ta/ and /da/ by the native listeners of Chubu region

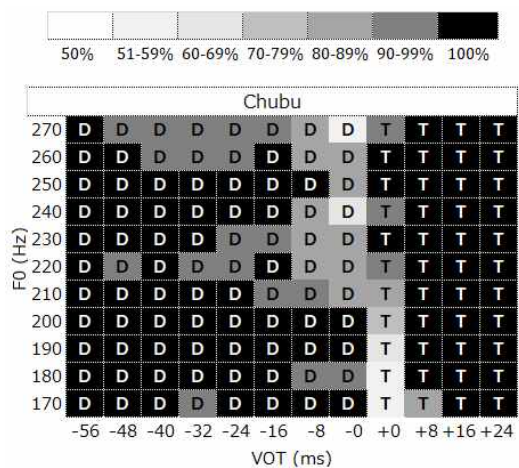


그림 11. 합성음 /ta/ /da/의 인식률(주부지방) 상세표시  
Figure 11. Identification rates of synthesized sounds /ta/ and /da/ by the native listeners of Chubu region (in detail)

그림 11에서 VOT가 0 ms인 경우 +0 ms는 F0가 낮을 때 다른 양 값의 VOT보다 무성음 지각이 감소하고 -0 ms는 F0가 높을 때 다른 음 값의 VOT보다 유성음 지각이 감소하는 것은 VOT에 관한 유효한 정보가 없어 유성/무성의 구별이 어려울 때는 F0를 이용하는 증거가 될 수 있을 것 같다. 다만 양 값의 VOT 영역 전체와 음 값의 VOT 영역 전체를 비교해 보면 양상이 조금 다르다. 음 값의 VOT 영역에서는 VOT에 상관없이 높은 F0에 대해 무성음으로 지각하는 예가 증가하나 양 값의 VOT 영역에서는 낮은 F0에 대해서 유성음의 지각이 거의 없다. 그림 11에서 보는 한, 낮은 F0일 때 유성음, 높은 F0일 때 무성음으로 지각한다기 보다는 높은 F0일 때 무성음으로 지각하는 것으로 보인다. 이처럼 VOT로 판단이 애매한 경우에 낮은 F0를 가지고 유성음으로 판단하기 보다는 높은 F0를 가지고 무성음으로 판단하는 예를 일본어는 Gao et al.(2019), 영어는 Whalen et al.(1990)에서 확인할 수 있다.

#### 4. 논의

##### 4.1. 일본어 어두 폐쇄음의 일차적 지각 변수는 VOT

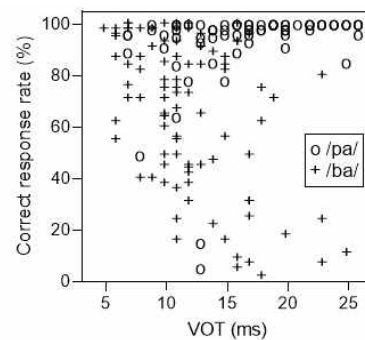
합성음을 사용한 실험 3에서 F0의 관여는 인정되나 그 정도는 극히 한정적이며 일본어 유성/무성의 지각은 전적으로 VOT에 의해 행해지는 것을 확인하였다. 여기서는 자연음을 사용한 실험 1에서도 F0가 거의 쓰이지 않았다는 것을, 관련한 선행연구와 비교하면서 밝혀보고자 한다.

자연음을 사용한 실험 1의 정답률은 도호쿠, 간토, 긴키, 주부의 4개 지역 모두 무성음은 평균 93%~95%, 유성음은 63%~69%로 특히 유성음에서 상당히 낮았다. 자연음의 VOT는 0~25 ms로 선택지가 두 개(유성음/무성음)밖에 없는 것을 고려하면 많은 청자가 0~25 ms의 VOT를 갖는 유성음을 무성음으로 선택한 것을 알 수 있다. 이러한 경향은 어느 지역이나 마찬가지로, VOT 사용에 있어서 산출에서는 지역에 따른 차이가 있지만 지각에서는 지역과 상관없이 양 값의 VOT일 때 무성음으로 지각하는 것으로 추정된다. 이는 동일 지역 내의 화자(=청자)를 조사한 Matsui(2014)의 결과와도 일치한다. Matsui는 Klatt speech synthesizer로 작성한 합성음을 이용한 결과이고 본고의 실험 1은 자연음을 이용한 결과로 자극음의 성질은 다르나 상기 외에도 결과에 유사한 점이 있으므로 간단히 언급한다.

Matsui(2014)는 긴키방언 화자(=청자) 29명을 대상으로 산출과 지각 실험을 실시하였다. 산출에서는 유성음의 VOT가 거의 음 값만을 갖는 화자(3명)와 그렇지 않은 화자(26명)로 나뉘었으나 지각에서는 어느 청자나 10~20 ms를 경계로 유성/무성을 인지하는 것으로 나타났다(그림 인용 생략). 구체적으로는 VOT가 10 ms이하이면 80% 이상이 유성음으로 지각하고 20 ms이상이면 80% 이상이 무성음으로 지각하였다. 자극음의 간격은 10 ms로 10~20 ms 사이의 결과는 알 수 없으나 이 10~20 ms가 유성/무성 구별의 경계치로 추정된다.

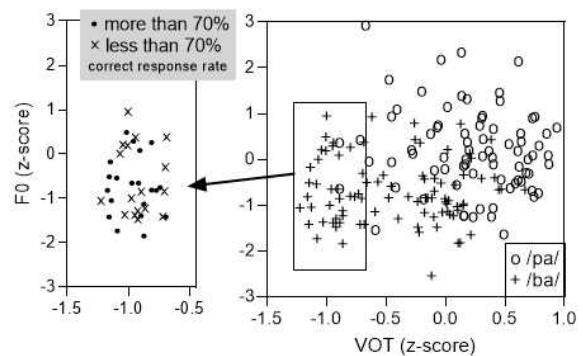
위의 Matsui(2014)의 결과를 본고의 실험 1과 비교해 보자. 그림 12에 실험 1의 VOT에 따른 정답률을 제시하였다. 실험 1은

VOT가 0~25 ms로 10~20 ms가 어떤 상태인지를 확인할 수 있다. 무성음은 거의가 90% 이상이지만 VOT가 10 ms 전후에서는 5개의 자극음(O표시)이 80%에 미치지 못하였다. 유성음은 Matsui의 결과와 유사하여 대략 10 ms 이하는 유성음의 정답률이 높고, 20 ms 이상에서는 유성음의 정답률이 매우 낮다(즉, 무성음으로 지각). 중간인 10~20 ms는 10 ms에 가까운 자극음이 많은 탓에 유성음의 정답률이 평균 60%이지만 자극음 수가 고르게 갖추어진다면 50%정도가 될 것으로 보인다. Matsui의 자극음은 합성음으로 /ta-/da/로 선택, 본고의 실험 1의 자극음은 자연음으로 /pa-/ba/로 선택한다는 차이는 있으나 치경음 /t-/d/와 양순음/p-/b/의 VOT 값이 매우 가까워서 비교하는 데에 큰 무리는 없을 것이다. 여기에서 중요한 것은 유성음이 VOT에 반응하여 인지되고 있다는 점이다.



VOT, voice onset time.

그림 12. VOT(ms)에 따른 평균 정답률(실험 1)  
Figure 12. Average correct response rates by VOT (Exp. 1)



VOT, voice onset time.

그림 13. VOT(z-score)와 F0(z-score)에 따른 평균 정답률(실험 1)  
Figure 13. Average correct response rates by VOT and F0 (Exp. 1)

실험 1에서 유성음이 F0에 거의 반응하지 않았다는 것을 그림 13에서 확인할 수 있다. 그림 13은 실험 1의 자극음을 VOT와 F0로 나타낸 것이다. 자극음에 남녀의 음성이 섞여 있어 F0를 개인별로 정규화하고 마찬가지로 VOT도 정규화하였다. 오른쪽 쪽 그림의 □안에 있는 유성음(+표시) 37개의 정답률은 2%에서 100%로 중앙치는 71%이다(□안의 O표시 무성음 3개는 제외). 정답률 70%를 기준으로 그보다 높은 음과 낮은 음으로 나누어



표시한 것이 왼쪽의 그림이다. 왼쪽 그림에서 보는 것처럼 낮은 F0에서 유성음의 정답률이 높아지고, 높은 F0에서 유성음의 정답률이 낮아지는 식의 F0와 정답률과의 상관을 전혀 찾아볼 수가 없다(Pearson  $r=-0.045$ ,  $p=0.7893$ ). □안에 있는 유성음 37개를 포함하는 유성음 전체( $n=91$ )를 보면 F0와 정답률이 약한 음의 상관을 보이거나 유의하지 않았다( $r=-0.1326$ ,  $p=0.2101$ ). 무성음 전체( $n=77$ )에 대해서는 약한 양의 상관( $r=0.2586$ ,  $p=0.0232$ )이 있는 것으로 나타났는데 이것은 실험 3의 결과에서 확인한 높은 F0일 때 무성음 지각이 증가하는 것과 모순되지 않는다.

정리하면, 일본어 자연음의 경우 VOT는 무성음과 유성음의 지각에 모두 관여하며 어두 폐쇄음 지각의 일차적 변수로 볼 수 있다. 후속모음 F0는 유성음 지각에는 관여하지 않고 무성음에서는 약한 상관을 보였다. 즉 극히 한정적이기는 하지만 높은 F0는 무성음의 지각 단서로 쓰였으나 반대로 낮은 F0는 유성음의 지각 단서로 쓰이지 않았다.

#### 4.2. 음향 변수 이외의 지각 단서

유성/무성의 구별에 VOT만으로 구별이 불충분한 상태에서 이용 가능성이 가장 높은 F0를 이용하지 않는다면 어떻게 유성/무성을 구별하여 이해하는 것일까. 실험 1에서 사용한 자연음은 무의미어이지만 일본어에서 단음절로 된 유성/무성 폐쇄음의 최소 대립쌍(minimal pair)은 ka(蚊 모기) vs ga(蛾 나방), kin(金 금) vs gin(銀 은), to(都 수도) vs do(度 도), pan(パン 빵) vs ban(晩 밤) 등 어렵지 않게 찾아볼 수 있다. 필자(서울방언 청자)가 텔레비전이나 모어화자와의 대화에서 들려오는 어두 폐쇄음을 VOT만을 단서로 하여 들으려고 하면 VOT가 양 값인 유성음의 경우 뜻을 모르고는 제대로 이해할 수가 없다. F0를 단서로 하여 들으려고 하면 피치 악센트가 있어 첫음절이 낮게 시작하는 무성음은 그림 9의 서울방언 청자처럼 높은 F0를 무성음, 낮은 F0를 유성음으로 지각하는 사람에게는 유성음인지 무성음인지 매우 혼란스럽다. Gao & Arai(2019)는 산출에서 어두 폐쇄음의 구별에 사용되는 F0 차이는 첫음절에만 나타나므로 피치 악센트와는 충돌하지 않는다고 하였으나 피치 악센트의 존재는 분절음의 지각에 F0 사용을 저지하는 요인이 될 수 있을 것 같다.

앞에서도 말했지만 VOT만으로 유성/구별이 안 되어도 아직 일상생활에서 이로 인한 의사소통에 문제가 있다고 하는 보고는 나와 있지 않다. 그렇다면 모어청자는 무엇을 단서로 유성/무성의 단어를 구별하고 있을까. F0보다는 효과가 작았지만 산출에서 어느 정도 관련이 확인된 모음의 음질(H1-H2)을 의심해 볼 수도 있을 것이다. 본고에서는 모음의 음질보다도 영향력이 큰 F0를 다루었으나 서론에서 말한 것처럼 산출과 지각에서 쓰이는 음향 파라미터가 다를 수 있고 산출에서는 효과가 작아도 지각에서는 효과가 클 수도 있으므로 확인해 볼 필요가 있었다.

음향 파라미터 이외의 변수로는 예를 들면 문맥을 생각할 수 있다. 우리말에서도 장단의 최소 대립쌍인 눈(snow)과 눈(eye)이 장단의 소멸로 인해 단독발화에서는 구별이 어렵지만 이로 인해 의사소통에 문제가 발생하는 일은 매우 드물다. 실제로 어

느 쪽으로도 해석될 수 있는 경우는 그리 많지 않아 대부분은 문맥으로 이해할 수 있기 때문이다.

Kurotsaki(2002)는 분절음이 애매한 경우에 문맥으로 이해하는 예를 소개하고 있다. 일본어의 음절말 비음 ‘ん’은 후속음과 동일한 조음점의 비음으로 발음되는데 후속음이 모음, 반모음, 마찰음(구강내 접촉 없이 조음되는 음)일 때는 변이음인 비모음으로 발음되는 것이 보통이다(구체적으로 어떤 비모음인가에 대해서는 연구자 간에 불일치가 있으므로 여기서는 언급하지 않는다). 모어 화자는 통상 의식하지 않고 발음하기 때문에 예를 들면 단독으로 발음한 ‘てんいん(店員 점원)’을 들려주고 쓰게 하면 비모음으로 발음되는 첫 번째 ん을 모음으로 대용한 오답이 나온다. Kurotsaki는 이를 바탕으로 첫 번째 ‘ん’의 자리에 모음이 들어간 ‘ていいん(定員 정원)’을 단독의 경우와 문맥이 들어간 경우를 설정하여 제시하였다. 문맥은 ‘알바는 편의점 정원(ていいん)이야’로 바르게는 ‘점원(てんいん)이야’이지만 일부러 틀린 단어 ‘정원(ていいん)이야’를 넣어 바르게 수정하여 듣는지를 실험하였다. ‘정원(ていいん)’이 단독으로 제시되었을 때 ‘정원’으로 응답한 것은 57%였으나 문맥으로 제시되었을 때는 0%로 피험자 14명 모두가 ‘알바는 편의점 점원이야’로 응답하였다. 다시 말해 ‘정원’이라는 잘못된 단어를 문맥에 맞게 ‘점원’으로 수정하여 듣는 것이 100%인 것이다.

음향 변수로만 보면 양 값의 VOT를 갖는 어두 유성 폐쇄음이 무성 폐쇄음과 구별이 어려운 상황에서 지각에 VOT 이외의 결정적인 다른 변수를 사용하지 않으면서도 의사소통에 혼란이 일어나지 않는 것은 유성/무성의 지각에 음향 파라미터가 아닌 문맥과 같은 요인이 관여하고 있을 가능성을 충분히 생각해 볼 만 하다. 실험 1에서 단독으로 제시한 유성 폐쇄음의 정답률은 평균 66%였다. 문맥이 폐쇄음의 유성/무성에 관여한다면 적절한 문맥 하에서 양 값의 VOT를 갖는 유성음의 정답률은 거의 100%가 될 것으로 예상된다. 또한 VOT가 무성 폐쇄음에 상당하는 음을 제시해도 문맥에 맞는 유성 폐쇄음으로 수정하여 답할 것으로 예상된다. 이를 증명하기 위해서는 악센트, 문맥 등을 적절히 갖춘 별도의 실험이 필요할 것이다.

#### 5. 결론

산출에서 일본어의 어두 폐쇄음이 VOT만으로 유성/무성의 구별이 어렵다는 것은 여러 선행연구에서 확인된 사실이다. 지역에 따라 정도의 차이는 있으나 후속모음의 F0 차이가 유성/무성의 구별에 관여하는 것도 선행연구에서 공통적으로 확인된 사실이다. 본고는 산출에 관여하는 음향 변수 VOT와 F0가 지각에서도 쓰이는지를 살펴보기 위하여 자연음과 합성음을 이용한 지각 실험을 실시하였다.

일본어 어두 폐쇄음의 일차적 지각 변수는 VOT로, 예상했던 F0의 사용은 극히 한정적인 것으로 확인되었다. 이는 지역과 상관없이 어느 지역에서나 같은 결과였는데 산출에 쓰이는 음향 변수와 지각에 쓰이는 음향 변수가 일치하지 않는 것을 보여준다. VOT만으로 카테고리 구별이 어려운 데에도 불구하고 F0를

사용하지 않는다면 또 다른 변수를 생각해 봐야 할 것이다. 다른 변수의 존재를 인정하지 않고는 일상생활에서 유성/무성으로 인한 혼란이 없는 것을 설명할 수 없기 때문이다. 본문에서는 모음의 음질(H1-H2) 외에도 문맥과 같은 음향 변수 이외의 변수가 관여하고 있을 가능성에 대해 논하였다. 전자의 증명은 본고의 합성음을 이용한 실험과 같이 무의미어의 단음절로도 가능하겠지만 후자를 증명하기 위해서는 문맥을 설정한 유의미어를 통해 밝혀져야 할 것이다.

실험 3에서는 주부지방 청자가 F0 사용이 극히 한정적인 것을 보여주었는데 이에 대해서는 재고의 여지가 있다. 주부지방은 산출에서 VOT를 주로 사용하고 F0 사용이 가장 소극적인 지역이다. 실험 3은 주부지방 이외의 청자에 대한 조사가 여의치 않아 현재 보류된 상태로 본고에서는 확인이 불가능했으나 후속모음 F0 차이가 지각에서 유효하지 않다는 결론을 내리기 위해서는 산출에서 F0가 VOT와 거의 동등하게 주요 변수로 작용하고 있는 도호쿠지방을 조사해야 할 것이다. 실험 3이 도호쿠지방에서도 같은 결과를 얻는다면 F0 차이가 유효한 지각 변수가 아니라고 결론지을 수 있을 것이다. 이에 대해서는 위의 모음의 음질 차이, 문맥 효과의 확인과 함께 금후의 과제로 삼겠다.

## References

- Abramson, A. S., & Lisker, L. (1970, January). Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 569-573). Prague, Czech.
- Abramson, A. S., & Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 25-33). Orlando, FL: Academic Press.
- Abramson, A. S., & Lisker, L. (1973). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics*, 1(1), 1-8.
- Abramson, A. S., & Whalen, D. H. (2017). Voice onset time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, 63, 75-86.
- Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer (version 6.0.22) [Computer program]. Retrieved from <http://www.praat.org/>
- Byun, H. G. (2021). Acoustic characteristics for Japanese stops in word-initial position: VOT and post-stop fo. *Onsei Kenkyū*, 25, 41-63.
- Gao, J., & Arai, T. (2019). Plosive (de-)voicing and f0 perturbations in Tokyo Japanese: Positional variation, cue enhancement, and contrast recovery. *Journal of Phonetics*, 77, 100932.
- Gao, J., Yun, J., & Arai, T. (2019, August). VOT-F0 coarticulation in Japanese: Production-biased or misparsing? *Proceedings of the 19th ICPHS* (pp. 210-214). Melbourne, Australia.
- Homma, Y. (1980). Voice onset time in Japanese stops. *Onsei Gakkai Kaiho*, 163, 7-9.
- Itoh, M., Sasanuma, S., Tatsumi, I. F., & Kobayashi, Y. (1979). Voice onset time characteristics of apraxia of speech. *Brain and Language*, 17(2), 193-210.
- Kang, J., & Kong, E. (2014). The role of f0 cue in distinguishing a laryngeal contrast of word-initial stops in Korean and Japanese. *Language and Linguistics*, 62, 75-100.
- Kim, M. (2004, October). Correlation between VOT and F0 in the perception of Korean stops and affricates. *Proceedings of the Interspeech 2004* (pp. 49-52). Jeju, Korea.
- Kobayashi, T. (1981). Nihongo no gotō haretsuon no VOT [VOT in Japanese word-initial stops]. *Gengo Bunka Kenkyūjo*, 7, 149-157.
- Kong, E. J. (2012). Perception of Korean stops with a three-way laryngeal contrast. *Phonetics and Speech Sciences*, 4(1), 13-20.
- Kong, E. J., Beckman, M. E., & Edwards, J. (2012). Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *Journal of Phonetics*, 40(6), 725-744.
- Kurosaki, N. (2002). Boin ni zensetsusuru hatsuon nitsuite [On the syllable-final nasal followed by a vowel]. *Kanagawa University Studies in Language*, 25, 11-22.
- Lee, H., Politzer-Ahles, S., & Jongman, A. (2013). Speakers of tonal and non-tonal Korean dialects use different cue weightings in the perception of the three-way laryngeal stop contrast. *Journal of Phonetics*, 41(2), 117-132.
- Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3), 384-422.
- Lisker, L., & Abramson, A. S. (1970, January). The voicing dimension: Some experiments on comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 563-567). Prague, Czechoslovakia.
- Matsui, M. (2014, September). Hensokutekina gōseion'o mochiita shiin yūseisei no chikakuni kansuru tegakari nitsusite [Perceptual cues to stop voicing with anomalous synthetic sounds]. *Proceedings of the 28th General Meeting of the Phonetic Society of Japan* (pp. 117-122). Tokyo, Japan.
- Ōtomo S., Sato, T., & Takakura, M. (1957). Nihongono musei haretsuon to yūsei haretsuon no sōi [On the difference between voiceless stop consonants and voiced stop consonants in Japanese sounds]. *The bulletin of the Phonetic Society of Japan*, 95, 4-9.
- Riney, T. J., Takagi, N., Ota, K., & Uchida, Y. (2007). The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics*, 35(3), 439-443.
- Sato, T. (1958). On the differences in time structures of voiced and unvoiced stop consonants. *The Journal of the Acoustical Society of Japan*, 14, 117-122.
- Shimizu, K. (1996). *A cross-language study of voicing contrasts of*

- stop consonants in Asian languages*. Tokyo, Japan: Seibido.
- Shimizu, K. (1999). Nichieigo niokeru heisashiin no yūseisei-museisei no onseiteki tokuchō [A study on phonetic characteristics of voicing of stop consonants in Japanese and English]. *Onsei Kenkyū*, 3(2), 4-10.
- Takada, M. (2004). Nihongo no gotō no yūsei shikei haretsuon /d/ niokeru +VOTka to sedaisa [The +VOT tendency in initial alveolar plosive /d/ in Japanese and speakers' age]. *Onsei Kenkyū*, 8(3), 57-66.
- Takada, M. (2006). Gotō yūsei haretsuon niokeru VOT no chi' ikisa to sedaisa: Tohoku kara Kanto no bunseki [Regional and generational variations of VOT in initial voiced plosives: An analysis of the Tohoku-Kanto area]. *Nihongo no Kenkyū*, 2(2), 34-45.
- Takada, M. (2011). *Nihongo gotō heisaon no kenkyū: VOT no kyō jiteki bunpu to tsūjiteki henka* [A study on the word-initial stops of Japanese: Synchronic distribution and diachronic change in VOT]. Tokyo, Japan: Kurosio.
- Takada, M., Kong, E. J., Yoneyama, K., & Beckman, M. E. (2016, October). Do pitch and voice quality cue word-initial “voicing” in Tōhoku Japanese? *The 24th Japanese/Korean Linguistics Conference*. Tachikawa, Japan.
- Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2013). Roles of voice onset time and F0 in stop consonant voicing perception: Effects of masking noise and low-pass filtering. *Journal of Speech, Language, and Hearing Research*, 56(4), 1097-1107.
- Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1990). Gradient effects of fundamental frequency on stop consonant voicing judgments. *Phonetica*, 47(1-2), 36-49.
- Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1993). F0 gives voicing information even with unambiguous voice onset times. *The Journal of the Acoustical Society of America*, 93(4), 2152-2159.

• 변희경 (Hi-Gyung Byun) 교신저자

국제교양대학 준교수

Akita International University

Yuwa, Akita-city, Akita 0101292, Japan

Email: byun@aiu.ac.jp

관심분야: 음성학, 한일음성 대조연구, 음성교육

## 모어청자에 의한 일본어 어두 폐쇄음의 지각\*

변희경

국제교양대학

### 국문초록

일본어의 어두 폐쇄음은 일차적으로 VOT로 구별되며 유성음은 음 값의 VOT, 무성음은 양 값의 VOT를 갖는 것으로 알려져 있다. 최근의 연구에 의하면 유성음이 양 값의 VOT를 갖는 어두 유성 폐쇄음의 무성화 현상이 전국적으로 젊은 층을 중심으로 진행되고 있는 것이 확인되었다. 또한 지역마다 정도의 차이는 있으나 어두 폐쇄음 구별에 VOT 이외에 후속 모음의 F0 차이가 유효한 것이 밝혀졌다. 본고는 산출에 쓰이는 음향 변수 VOT와 F0가 지각 변수로도 유효한지를 검토한 것이다. 4개 지역의 대학생 55명을 대상으로 자연음과 합성음을 이용한 세 가지 지각 실험을 실시한 결과는 어느 지역에서나 어두 폐쇄음의 일차적 지각 변수로 기능하는 것은 VOT이며 예상했던 F0의 사용은 극히 한정적인 것으로 나타났다. F0는 VOT가 기능하지 않을 때에 높은 F0를 무성음으로 지각하는 경우는 있으나 반대로 낮은 F0를 유성음으로 지각하는 경우는 거의 보이지 않았다. 이러한 결과는 유성/무성의 구별에 관여하는 산출의 음향 변수와 지각 변수가 일치하지 않는 것을 뜻하며, 나아가 일차적 변수인 VOT의 구별이 어려운 상황에서도 F0의 적극적인 사용 없이 일상생활에서 유성/무성의 구별에 혼란이 없는 것은 VOT와 F0 이외의 다른 요인이 관여하고 있음을 시사한다. 본고에서는 산출에서 어느 정도 관여가 인정된 모음의 음질(H1-H2)이나 음향 변수가 아닌 문맥이 관여하고 있을 가능성에 대해 논하였으나 이에 대한 검증은 앞으로의 과제로 남아 있다.

**핵심어:** 일본어, 어두 폐쇄음, 지각 특성, VOT, 후속모음F0

### 참고문헌

강지은, 공은정 (2014). 한국어와 일본어의 어두 폐쇄음 후두자질  
구분에 있어 음향단서 f0의 역할. *언어와 언어학*, 62, 75-100.

\* 본 연구는 JSPS KAKENHI의 지원을 받아 수행되었습니다(과제번호: JP19K00634, JP20K00642).