



The role of voice onset time (VOT) and post-stop fundamental frequency (F0) in the perception of Tohoku Japanese stops*

Hi-Gyung Byun**

Faculty of International Liberal Arts, Akita International University, Akita, Japan

Abstract

Tohoku Japanese is known to have voiced stops without pre-voicing in word-initial position, whereas traditional or conservative Japanese has voiced stops with pre-voicing in the same position. One problem with this devoicing of voiced stops is that it affects the distinction between voiced and voiceless stops because their voice onset time (VOT) values overlap. Previous studies have confirmed that Tohoku speakers use post-stop fundamental frequency (F0) as an acoustic cue along with VOT to avoid overlap. However, the role of post-stop F0 as a perceptual cue in this region has barely been investigated. Therefore, this study explored the role of post-stop F0 in stop voicing perception along with VOT. Several perception tests were conducted using resynthesized stimuli, which were manipulated along a VOT continuum orthogonal to an F0 continuum. The results showed no significant regional difference (Tohoku vs. Chubu) for nonsense words (/ta-da/). However, for meaningful words (/pari/ 'Paris' vs. /bari/ 'Bali,' /piza/ 'pizza' vs. /biza/ 'visa'), a significant word effect was found, and it was confirmed that some listeners utilized the post-stop F0 more consistently and steadily than others. Based on these results, we discuss innovative listeners who may lead the change in the perception of stop voicing.

Keywords: Tohoku Japanese, word-initial stop voicing, perceptual cues, voice onset time (VOT), post-stop fundamental frequency (F0)

1. 서론

전통적인 일본어의 폐쇄음은 어두에서 파열 전에 성대 진동 (pre-voicing)을 동반하는 유성음/b d g/와 파열 후에 약간의 기음 (aspiration)을 동반하는 무성음/p t k/로 이분된다. 그러나 젊은 세대에서는 유성음이 파열 전에 성대 진동을 동반하지 않는 음으로 생성되는 경우가 많아 전국적으로 어두 유성폐쇄음의 무

성화 현상이 진행 중인 것이 알려져 있다(Takada, 2004; Takada, 2011). Takada(2011)에 의하면 유성폐쇄음의 무성화가 보이기 시작한 것은 1920년대 이후에 출생한 화자로, 일본 북쪽 끝에 위치한 홋카이도(Hokkaido)와 그 아래에 있는 도호쿠(Tohoku) 지방을 제외한 전국에서 관찰된다. 이러한 무성화 현상은 지역마다 진행 속도가 조금씩 다른데 그 중 나고야를 포함한 주부(Chubu) 지방은 젊은 세대도 전통적인 유성음(pre-voicing)을 보

* This work was in part supported by JSPS KAKENHI (Grant Numbers: JP19K00634, JP20K00642).

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

Received 10 February 2023; Revised 7 March 2023; Accepted 20 March 2023

© Copyright 2023 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.

유하고 있는 화자가 많아 유성폐쇄음의 무성화에 대해 가장 보수적인 지역이라 할 수 있다(Byun, 2021a). 한편 홋카이도와 도호쿠지방은 어느 세대나 본래 유성폐쇄음이 파열 전에 성대 진동을 동반하지 않는 음(short-lag)으로 생성되어 다른 지역들과 큰 대조를 이룬다. 1990년대 출생자가 발음한 1음절 단독 발화의 경우, 어두 유성폐쇄음의 무성화율(유성음이 성대 진동을 동반하지 않는 비율)은 주부지방이 26%, 도호쿠지방이 83%(Byun, 2021a)로, 이 두 지역이 전국 본포의 양극을 형성한다. 이 외 지역의 무성화율은 이들 중간에 위치하며 예를 들어 동경을 포함한 간토(Kanto) 지방의 무성화율은 58% 정도이다(Byun, 2018; Takada, 2011).

서울말의 연음과 격음이 voice onset time(VOT) 차이가 모호해지면서 변별 기능이 후속 모음의 기본주파수(fundamental frequency, F0) 차이로 바뀐 것처럼 일본의 어두 폐쇄음에도 유사한 현상이 진행 중이다. 유성음의 무성화로 인해 유성음의 VOT가 음값(pre-voicing)이 아닌 양값(short-lag)으로 나타나는 세대에서는 본래 중간 정도의 기음(aspiration)을 갖는 무성음과 중복하게 되어 VOT로 유성음과 무성음의 구별이 어려워졌고 이러한 세대에서는 F0가 변별 기능의 일부를 대신하는 것이 확인되었다(Byun, 2018; Byun, 2021a; Gao & Arai, 2019; Kong et al., 2012). 다만 F0 사용에는 지역차가 있어서 유성음의 무성화율이 낮고 VOT 역할이 큰 주부지방에서는 F0가 거의 사용되지 않는 데에 반해 무성화율이 높은 도호쿠지방에서는 F0가 VOT와 거의 같은 정도로 폐쇄음 구별에 중요한 역할을 하는 것이 보고되어 있다(Byun, 2021a).

일본어 어두 폐쇄음의 변별 기능의 일부가 VOT에서 F0로 바뀐 것과 관련하여, 도호쿠지방의 경우, 1920년대 출생자의 발화에는 없었던 유성음과 무성음 사이의 F0 차이가 1980년대 출생자의 발화에서는 유의하게 기능하고 있는 것이 확인되었다(Takada, 2021). 이에 대해 Takada는 VOT만으로는 부족한 변별 기능을 보완하기 위하여 기존의 음향특성이 아닌 새로운 음향 특성으로 F0가 쓰이게 되었다고 해석하였다. 위에서 언급한 것처럼 도호쿠지방의 화자는 윗세대에서도 유성폐쇄음이 성대 진동을 동반하지 않는 것이 보통으로 본래 유성음과 무성음의 VOT가 매우 근접해 있었으나 세대가 내려올수록 (유성음이 아닌) 무성음의 양값의 VOT가 조금씩 줄어든 결과(Takada, 2011), 아랫세대에서 VOT의 중복이 많아졌고 이를 해소하기 위해 F0 차이가 발생한 것으로 보인다.

이렇게 생성에 있어서는 VOT의 기능이 약화된 지역에서 이를 대신할 음향특성으로 F0가 관여하고 있는 것이 확인되었으나 지각에 있어서는 F0가 유효한지에 대해서는 충분한 연구가 이루어지지 않은 상태이다. 따라서 본 연구에서는 후속 모음 F0가 어두의 폐쇄음 구별에 관여하는 지각특성으로 유효한지를 검토하였다.

본고의 구성은 다음과 같다. 먼저 2.에서 선행연구를 살펴보고 이를 바탕으로 실시한 지각실험에 대해 3.에서 상술한다. 4.에서 결과를 제시하고 5.에서 지각의 개인차와 음변화를 견인하는 혁신 청자에 대해 논의한 후 6.에서 결론으로 끝맺는다.

2. 선행연구

일본어 어두 폐쇄음의 지각특성으로 VOT와 F0를 동시에 검토한 선행연구로 Byun(2021b)과 Gao et al.(2019)이 있다. 전자는 동경에 거주하는 20-30대(출신지역 불명), 후자는 (3개의 실험 중 VOT의 음값과 양값을 모두 포함하는 실험3의 경우) 주부지방의 10-20대 젊은 층을 피험자로 하였다. 모두 유성폐쇄음의 무성화를 경험하고 있는 세대이다. 두 연구 모두 완전 합성음이 아닌 자연음을 원음으로 하여 재합성한 자극음을 사용하였고 실험 결과도 거의 일치하였다. 즉, VOT의 카테고리 경계는 0 ms 전후이고 어두 폐쇄음의 변별에 쓰이는 일차적 지각특성은 VOT이지만 VOT로 판단이 어려운 경우에는 F0가 부차적으로 이용되었다. 이 때에 낮은 F0를 유성음으로 판단하는 것이 아니고 높은 F0를 무성음으로 판단하는 경향이 있는 것도 일치하였다. 순서대로 자세히 살펴보겠다.

Gao et al.(2019)의 자극음은 2음절 유의미어(HL음조: /pasu/ ‘패스’ vs. /basu/ ‘버스’, LH음조: /teki/ ‘적’ vs. /deki/ ‘결과, 완성된 상태’)로 VOT는 10-15 ms 간격으로 8단계(-60 ms, -45 ms, -30 ms, -15 ms, +10 ms, +20 ms, +30 ms, +40 ms), F0는 음조별로 첫음절의 H음조(111-154 Hz)와 L음조(114-133 Hz)를 등간격으로 각각 6단계로 나눈 것을 조합한 96음이다. 이것을 청자가 3회 또는 4회 반복 청취하였다. 청취 결과를 보면 VOT의 카테고리 경계는 -15 ms와 +10 ms 사이로(0 ms의 설정은 없고 음값과 양값 사이는 25 ms 간격), H음조는 0 ms 부근에서, L음조는 +10 ms에서 F0가 높을 때 무성음 판정이 우세하였다. VOT의 경계치는 0 ms 부근이 25 ms 간격으로 범위가 넓어서 상세를 확인할 수 없는 것이 아쉽다.

Byun(2021b)의 자극음은 1음절 무의미어 /ta/, /da/로 VOT는 -56 ms에서 +24 ms까지를 8 ms 간격으로 11단계(0 ms는 유성음이 원음인 경우 -0 ms, 무성음이 원음인 경우 +0 ms로 설정), F0는 170 Hz에서 270 Hz까지를 10 Hz 간격 11단계로 조작한 132음이다. 주부지방 출신의 청자 13명이 2회 반복 청취하였다. 청취 결과는 다소나마 -0 ms에서 높은 F0일 때 무성음 판정이 증가하고 +0 ms일 때 낮은 F0에서 유성음 판정이 증가하는 경향을 보였다. 다만 VOT가 양값인 자극음에 대해서는 F0에 상관없이 무성음으로만 판정하나 VOT가 음값인 자극음에 대해서는 F0가 높은 경우에 유성음 판정이 상대적으로 증가하는 것을 바탕으로 낮은 F0보다는 높은 F0에 반응한다고 해석하였다.

Byun(2021b)은 주부지방의 결과로 저자가 지적하고 있는 것처럼 이것만으로 일본어 어두 유성음 전체의 지각특성을 설명하는 것은 적절하지 않다. 왜냐하면 주부지방은 본래 생성에서 F0를 거의 사용하지 않기 때문에 지각에서 F0 효과가 거의 없는 것이 어찌면 당연할 수도 있기 때문이다. 서론에서도 언급했듯이 주부지방과 도호쿠지방은 일본어 어두폐쇄음의 VOT 실험에 있어서 양 끝점에 위치한다. 일본어 폐쇄음 전체의 지각특성을 설명하기 위해서는 현재 VOT가 변화 중인 지역을 조사하는 것도 중요하겠지만 그에 앞서 양 끝점을 이루는 지역의 특징을 밝히는 것이 선행되어야 할 것이다. 변화 중인 지역이 어느 정

도 변화했는지를 판단하는 기준이 되기 때문이다. 본 연구에서는 도호쿠지방의 청자를 연구 대상으로 하되 주부지방의 결과와도 비교할 수 있도록 하였다.

본고의 목적은 도호쿠지방 청자를 대상으로 지각특성으로서 VOT와 F0의 역할을 밝히는 것이다. 무의미어와 유의미어를 사용한 지각실험을 통해 동일 조건의 선행연구(주부지방)의 결과와 비교하여 유성음, 무성음 지각에 지역 차이가 있는지를 살펴보고 유의미어의 경우 VOT와 F0 사용에 어휘 차이가 있는지를 밝혔다. 또한 F0 사용에 보이는 개인차와 이 개인차가 갖는 과급효과에 대해 음변화와 관련하여 논의하였다.

3. 지각실험

세 종류의 자극음을 사용하여 실험1에서 실험3까지 세 가지 지각실험을 실시하였다. 실험1은 F0 사용에 있어서 주부지방과 도호쿠지방 사이에 차이가 있는지를 보기 위한 것으로, 자극음은 주부지방 청자를 대상으로 한 Byun(2021b)에서 1음절의 무의미어 /ta-da/를 빌려왔다. 실험2와 실험3은 도호쿠지방 청자가 유의미어에서 폐쇄음 변별에 F0를 어느 정도까지 사용하는지를 보기 위한 것으로, 자극음으로 2음절의 유의미어 /pari-bari/ (실험2)와 /piza-biza/ (실험3)를 사용하였다. 같은 자음, 다른 어휘로 동종의 실험(실험2, 실험3)을 실시한 이유는 한국어 폐쇄음 발화에서 VOT의 기능이 F0로 이행하는 데에 어휘 차이가 있는 것이 시사되었기 때문에(Bang et al., 2015; Bang et al., 2018), 발화와 지각의 차이는 있으나 일본어에서도 어휘 차이가 있는지를 보기 위해 설정하였다. 다만 여러 단어를 사용한 장시간에 걸친 지각실험은 참가자에게 부담이 되므로 참가자의 부담 경감을 위해 어휘 수는 2개의 최소대립쌍으로 한정하였다.

폐쇄음의 종류가 실험1은 치경음, 실험2, 3은 양순음으로 조음점이 다른데, 일본어의 치경음과 양순음 사이에는 VOT 차이가 거의 없기 때문에(Riney et al., 2007; Takada, 2011) 지각실험의 단조로움을 피하기 위하여 다른 종류의 자음을 선택하였다. 같은 이유로 실험1과 실험2, 3의 원음의 성별을 달리하였다.

3.1. 자극음

3.1.1. 실험1 무의미어 /ta-da/

Byun(2021b)의 실험3에 사용된 자극음 /ta/와 /da/를 그대로 사용하였다. 여성이 단독 발화한 원음을 Praat(Boersma & Weenink, 2019, version 6.2.22)의 Manipulate 기능을 이용하여 VOT와 F0를 재합성한 것으로 합성 방법과 순서는 다음 절에서 설명하는 유의미어와 같다. 다른 점은 완성된 자극음이 단음절로 피치 변동이 거의 없어 로봇음과 같은 부자연스러움을 주기 때문에 F0를 시작점(onset)에서 끝점(offset)까지 일률적으로 5% 낮추었다. 5% 하강폭을 포함하여 상세한 것은 Byun(2021b)에 기술되어 있으므로 여기서는 생략한다.

VOT의 범위는 8 ms 간격으로 -56 ms에서 +24 ms까지 있지만 Byun(2021b)의 결과에서 -40 ms보다 작은 값에 큰 변화가 없

었기 때문에 참가자의 부담을 줄이기 위해 -56 ms와 -48 ms를 제외한 9단계(-40 ms, -32 ms, -24 ms, -16 ms, -8 ms, -0 ms, +0 ms, +8 ms, +16 ms, +24 ms)를 사용하였다(단, 0 ms는 원음의 VOT가 음값인 -0 ms와 원음의 VOT가 양값인 +0 ms를 포함).

F0는 170 Hz에서 270 Hz까지 10 Hz 간격의 11단계를 그대로 사용하였다. 따라서 자극음 수는 110 토큰으로 실험1에서는 이것을 2번 반복하여 한 사람당 총 220 토큰의 응답이 얻어졌다.

3.1.2. 실험2 유의미어 /pari-bari/

남성 성우가 단독 발화한 HL음조의 2음절어 /pari/ ‘파리’와 /bari/ ‘바리’이다. 현대 일본어에서 어두에 오는 /p/ 음은 차용어나 의성어, 의태어에만 쓰이는데 2음절어의 최소대립쌍으로 하다 보니 /b/음도 차용어가 되었다. 실험3도 같은 이유에서 차용어의 최소대립쌍이 되었는데, 네 단어 모두 일상적으로 쓰이는 단어로 음소 대립을 보는 자극음으로 문제가 없을 것으로 판단하였다. 원음 /pari/와 /bari/의 첫 자음 /p b/의 VOT는 각각 13 ms, -67 ms, 모음 /a/의 onset(시작점)은 148 Hz, 128 Hz였다. 첫음절 피치의 최고점은 각각 174 Hz, 171 Hz였다. /p b/이외의 /ari/의 길이와 피치는 둘 중 하나를 바꾸어 넣는 식으로 하여 완전히 동일한 음으로 통일하였다. 폐쇄음의 파열 순간의 강도(intensity)나 VOT 구간의 강도가 유무성 판정에 영향을 줄 수 있다는 지적이 있으나(Gao et al., 2019; Tamura et al., 2018), /p b/의 해당 부분의 강도에 큰 차이가 없었기 때문에 그대로 사용하였다.

VOT는 /ta-da/의 범위를 참고로 하여 8 ms 간격으로 -40 ms에서 +32 ms까지 11단계(상기와 마찬가지로 0 ms는 원음에 따라 -0 ms와 +0 ms의 둘을 포함), F0는 발화자의 피치폭을 참고로 하여 10 Hz 간격으로 90 Hz에서 180 Hz까지 10단계로 설정하였다. Praat(Boersma & Weenink, 2022, version 6.2.23)의 Manipulate 기능을 이용하여 먼저 VOT를 11단계로 조작한 후에 각각의 VOT 값에 대해 F0를 10단계로 조작하였다. F0는 시작점(onset: 모음의 주기적 파형이 시작되는 지점), 첫 번째 음절의 최고점, 두 번째 음절의 시작점, 중간점, 끝점을 지정한 후 onset을 변조하였다(첫 번째 음절의 끝점은 두 번째 음절의 시작점과 가까워서 전체에 영향을 주지 않으므로 미지정, 최고점 이후의 피치는 동일). 자극음의 총수는 110 토큰으로 2번 반복하여 한 사람당 총 220 토큰의 응답이 얻어졌다.

3.1.3. 실험3 유의미어 /piza-biza/

또 다른 유의미어는 같은 남성 성우가 단독 발화한 HL음조의 2음절어 /piza/ ‘피자’와 /biza/ ‘비자’이다. 원음의 첫 자음 /p b/의 VOT는 각각 16 ms, -91 ms, 모음 /i/의 onset은 160 Hz, 135 Hz였다. 첫 음절 피치의 최고점은 각각 182 Hz, 176 Hz였다. 실험2의 /pari-bari/의 경우와 마찬가지로 /piza-biza/에서 /p b/이외의 /iza/의 길이와 피치는 둘 중 하나를 바꾸어 넣는 식으로 하여 완전히 동일한 음으로 통일하였다. VOT와 F0의 조작은 실험2와 동일하며, 110 토큰을 2번 반복하여 한 사람당 총 220 토큰의 응답이 얻어졌다.

3.2. 청자

청자는 도호쿠지방에서 태어나(1998-2004년 출생) 18세까지를 같은 지역에서 생활한 원어민 13명(남성 1명, 여성 12명)으로 2022년 11월 실험 당시 대학생 또는 대학원생이었다(참가자의 성별이 많이 치우쳐 있어 성차에 대해서는 결과에서 언급하지 않겠다). 참가자는 실험 전에 실험에 관한 구두 설명을 듣고 자유의사에 의해 실험에 참가한다는 동의서에 자필 서명하였으며 실험 후에 소정의 사례를 받았다.

3.3. 수순

실험은 대학 내의 조용한 연구실에서 개인용 컴퓨터를 사용하여 Praat상에서 ExperimentMFC기능을 이용하여 실시되었다. 각 실험의 자극음은 개인마다 랜덤으로 제시되었다(randomize=<PermuteBalancedNoDoublets>). 참가자는 컴퓨터 화면에 쓰여 있는 지시문을 보면서 구두로도 같은 설명을 들었다. 실험1의 경우, 화면에는 일본어로 다음과 같은 지시문이 표시되었다. ‘협조해 주셔서 감사합니다. 이제부터 음성을 들을 겁니다. 들은 음성과 가장 가깝다고 생각되는 것을 [た] [だ] 중에서 고르세요. 어느 쪽도 아닌 것처럼 들려도 하나를 고르세요. 클릭과 동시에 다음 문제로 넘어갑니다. 음성을 다시 듣고 싶을 때에는 [다시 한 번]을 클릭하세요. 두 번까지 다시 들을 수 있습니다. 잘못 선택해서 다시 선택하고 싶을 때에는 [하나 앞으로]를 클릭해서 다시 하세요. 한 문제가 끝나면 화면 왼쪽 위에 있는 숫자가 바뀝니다(전체 220문제). 30문제마다 짧은 휴식이 있습니다. 그럼 클릭해서 시작하세요’. 화면에 표시되는 선택지는 일본어 가나로 표기하였고, 지시문에는 없으나 구두로 합성된 음이 포함되어 있어 부자연스럽게 들리는 것도 있으나 깊이 생각하지 말고 직관적으로 응답하도록 지시하였다. 참가자는 구두 설명 후에 노이즈 캔슬 기능이 있는 헤드폰(Sony WH-1000XM3, Tokyo, Japan)을 착용하고 실험을 개시하였다.

실험은 실험1 연습→실험1(/ta-da/)→실험2 연습→실험2(/pari-bari/)→실험3 연습→실험3(/piza-biza/)의 순서로 진행되었다(실험2 후에 동일 자극음, 실험3 후에 동일 자극음의 일부를 사용하여 유사한 실험을 하였으나 여기서는 언급하지 않겠다). 실험과 실험 사이에는 반드시 헤드폰을 벗고 휴식하도록 하였다. 본실험 화면에는 30문제마다 휴식을 유도하는 메시지를 넣었으나 휴식을 강요하지는 않았다. 실험 시간은 한 사람당 1시간을 넘지 않았다.

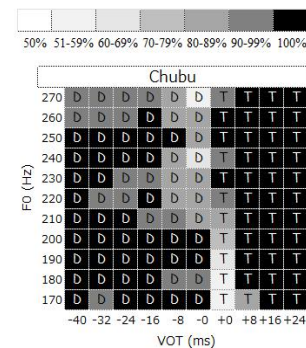
3.4. 통계

통계 분석은 R(R Core Team, 2022, version 4.2.2) 상에서 lme4 (Bates et al., 2015) 패키지(version 1.1-31)의 glmer 함수를 사용하여 이항 로지스틱 회귀분석을 실시하였다. 종속변인은 무성음과 유성음, 독립변인(고정효과)은 VOT와 F0 외에, 실험에 따라서 지역(도호쿠 vs. 주부) 또는 어휘(/pari-bari/ vs. /piza-biza/)이다. 의미(무의미어 vs. 유의미어)는 VOT와 F0의 범위가 다르고 F0값이 남녀 간에 달라서 직접 비교하기 어려워 생각하였다. 임의효과는 청자와 자극음이다.

4. 결과

4.1. 실험 1 무의미어 /ta-da/

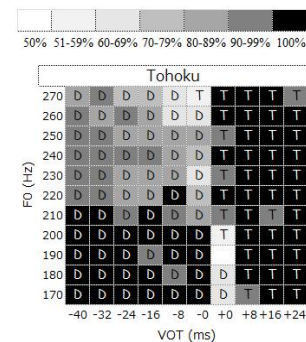
그림 1, 그림 2를 같이 보겠다. 그림 1은 주부지방의 결과로 비교를 위해 Byun(2021b, 그림11)에서 가져왔다. 그림 2는 실험 1의 도호쿠지방의 결과이다. VOT가 음값(-8 ms 이하)인 자극음을 먼저 보면, 두 지역 모두 유성음(D) 판정이 우세하였다. 다만 높은 F0에 대해서 주부지방은 유성음의 응답률(동정률)의 정도를 나타내는 색의 농담이 일정하게 변화하지 않고 F0에 따라서 진하다가 연하다가 하는 식으로 다소 무질서한 반응을 보이는 반면, 도호쿠지방은 대략 220 Hz 이상에서 VOT값에 따라 응답률의 농담이 점진적으로 변화하여 비교적 정연하게 반응하고 있는 것을 알 수 있다. 즉 VOT의 음값이 클수록 농담이 진하고 작을수록 연해져 0 ms에 가까울수록 유성음 판정이 감소(=무성음 판정이 증가)하였다. 하지만 무성음 판정이 50%를 넘지는 않아 최종판정은 유성음(D)로 남아 있다. 이에 반해 VOT가 양값(+8 ms 이상)인 자극음에서는 두 지방 모두 낮은 F0에 대해서 유성음으로 지각한 예는 거의 없다. 이러한 결과는 선행연구에서 언급한 낮은 F0를 유성음으로 판단하는 것이 아니고 높은 F0를 무성음으로 판단한다는 경향과 일치한다.



VOT, voice onset time; F0, fundamental frequency.

그림 1. 주부지방의 /ta-da/ 결과[Byun(2021b, 그림11, 일부 수정), Adapted from Byun et al.(2021b) with CC-BY-NC]

Figure 1. Result of /ta-da/ for Chubu listeners [Byun (2021b, Figure 11, excluding the VOT values of -56 ms and -48 ms), Adapted from Byun et al. (2021b) with CC-BY-NC]



VOT, voice onset time; F0, fundamental frequency.

그림 2. 도호쿠지방의 /ta-da/ 결과(실험1)

Figure 2. Result of /ta-da/ for Tohoku listeners (Exp. 1)

± 0 ms에서는 도호쿠지방에서 -0 ms일 때 높은 F0의 자극을 무성음, $+0$ ms일 때 낮은 F0의 자극을 유성음으로 판정하고 있는데, 이것은 그렇지 않은 주부지방보다 F0가 더 적극적으로 쓰인 것으로 볼 수 있겠다. 다만 표 1에서 보는 것처럼 유성음과 무성음 전체로 보면 지역 효과는 유의하지 않았다($p=0.396$).

표 1은 각 독립변인(VOT, F0, 지역)이 종속변인(무성음, 유성음)에 미치는 상대적 영향력을 보여준다. 계수(β)의 절대값이 클수록 유성음, 무성음 판정에 영향력이 큰 것을 의미하며 여기서는 VOT의 계수가 F0보다 크므로 VOT의 영향력이 크다고 할 수 있다. 다만 F0는 상대적인 영향력은 작아도 폐쇄음 구별에 유의하게 작용하고 있다. 이에 반해 지역 차이(주부, 도호쿠)의 영향력은 크지만 이미 말했듯이 통계적으로는 유의하지 않다. 표 1은 VOT와 F0를 같이 취급했을 때 두 변수를 나누어 지역과의 상호작용을 보면 상황은 조금 달라진다.

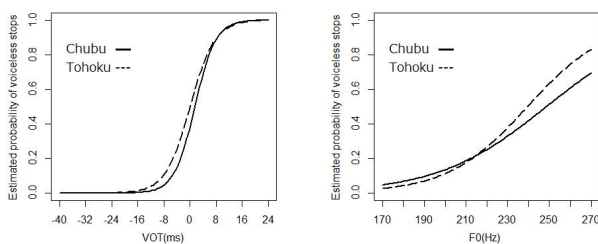
그림 3은 VOT와 F0를 지역별로 나누어 본 것이다. 회귀곡선의 기울기는 표 1의 계수에 상당하는 것으로, 경사가 급할수록 폐쇄음 판정에 영향력이 크고 경사가 완만할수록 영향력이 작은 것은 나타난다. 두 지역 모두 VOT의 기울기가 F0의 기울기보다 경사가 급하여 어느 지역이나 VOT의 영향력이 F0보다 크다고 할 수 있다. VOT와 F0 모두 지역과의 상호작용(VOT \times 지역, F0 \times 지역)이 유의하여(VOT: $\beta=-0.050$, $p<0.001$, F0: $\beta=0.217$, $p<0.0001$), F0의 경우, 주부지방보다 도호쿠지방에서 F0값이 클수록 무성음일 확률이 유의하게 높았다.

표 1. 그림 1과 그림 2의 일반화 혼합 모형 결과
Table 1. The summary of the mixed effects models for Figure 1&2

	β	SE	z-value	p-value
(Intercept)	-0.616	0.490	-1.256	0.209
VOT	0.284	0.023	12.250	<2e-16***
F0	0.055	0.011	4.669	3.03e-06***
AreaTohoku	0.407	0.479	0.849	0.396

*** $p<0.001$.

VOT, voice onset time; F0, fundamental frequency.



VOT, voice onset time; F0, fundamental frequency.

그림 3. 지역별 VOT와 F0의 회귀곡선
Figure 3. Regression curves of VOT and F0 for each region

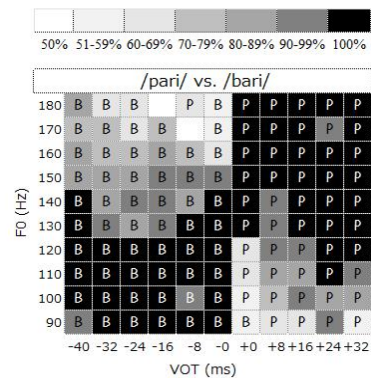
F0의 역할에 대해서 표 1과 그림 3의 결과가 모순되는 것처럼 보일 수 있으나 그렇지 않다. 표 1은 F0 외에 VOT의 영향이 있고 그림 3(오른쪽)은 VOT를 배제하고 F0만을 봤을 때를 예측한 것으로, F0만으로 보면 지역 차이가 인정되나 F0와 VOT와 같이 보면 (VOT의 영향력이 커서) 지역 차이가 유의하지 않게 나

온 것이다. 이것은 도호쿠지방에서 F0의 역할은 인정되나 지역 차이를 발생시킬 정도의 영향력은 갖지 못한다는 것을 뜻한다.

4.2. 실험 2 유의미어 /pari-bari/

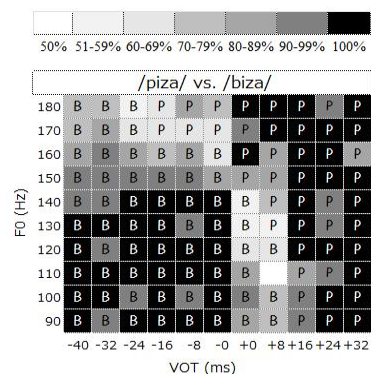
그림 4는 유의미어 /pari-bari/의 결과이다. 앞의 그림 2와 비교해 보면 그림 2의 무의미어의 경우보다 VOT가 음값일 때 높은 F0에서 무성음 판정이 다소 증가한 것을 알 수 있다(즉, 유성음 판정이 감소, 응답률의 농담이 더 연해짐). 다만 50%를 넘지는 않아서 최종판정은 유성음(B)로 남아 있다(180 Hz, -8 ms의 칸은 50%를 넘어 무성음(P), 빈칸은 유성, 무성의 판정이 각각 50%).

주목할 만한 것은 그림 2보다 VOT가 양값일 때 낮은 F0에서 유성음 판정이 증가했다는 것이다(즉, 무성음 판정이 감소, 응답률의 농담이 연해짐). 전체적으로 양값의 VOT와 높은 F0일 때 무성음, 음값의 VOT와 낮은 F0일 때 유성음 판정이 우세하여 무의미어의 결과보다 폐쇄음 지각에 F0가 보다 적극적으로 사용된 것을 확인할 수 있다. 높은 F0만이 아니라 낮은 F0에도 반응하는 것은 다음의 /piza-biza/에서도 분명히 나타났다.



VOT, voice onset time; F0, fundamental frequency.

그림 4. /pari-bari/의 결과(실험2)
Figure 4. Result of /pari-bari/ for Tohoku listeners (Exp.2)



VOT, voice onset time; F0, fundamental frequency.

그림 5. /piza-biza/의 결과(실험3)
Figure 5. Result of /piza-biza/ for Tohoku listeners (Exp.3)

4.3. 실험 3 유의미어 /piza-biza/

그림 5는 /piza-biza/의 결과이다. 그림 4의 /pari-bari/보다 F0 사용이 보다 적극적인 것을 볼 수 있다. VOT가 지각에 큰 역할을 하지 못하는 ± 0 ms 부근에서 높은 F0일 때 무성음으로 듣고 낮은 F0일 때 유성음으로 듣는 것은 그림 4와 같은 경향이지만, ± 0 ms만이 아니라 VOT가 음값(-16 ms, -8 ms)인 경우에도 높은 F0에서는 무성음 판정이 우세하고, VOT가 양값(+8 ms)인 경우에도 낮은 F0에 대해서 유성음 판정이 우세하다. +0 ms에서 유성음 판정이 많은 것도 그림 4와 다른 점이다. 다만 이러한 경향은 모든 청자에게서 고르게 보이는 것이 아니고 지각 패턴에는 개인차가 있어서 F0를 적극적으로 사용하는 청자가 있는가 하면 그렇지 않은 청자가 있다. 이러한 지각의 개인차에 대해서는 5. 논의에서 다시 언급하겠다. /pari-bari/와 /piza-biza/는 어휘 차이가 통계적으로 유의하여($\beta = -0.591$, $p < 0.001$) 유성음, 무성음 판정 전체에 있어서 /piza-biza/가 /pari-bari/보다 유성음일 확률(즉 /pari-bari/가 무성음일 확률)이 유의하게 높았다.

5. 논의

5.1. 일본어 어두 폐쇄음의 지각특성

도호쿠지방과 주부지방의 청자는 모두 어두 폐쇄음 지각에 VOT만이 아니고 F0를 사용하고 있다. 다만 무의미어의 경우 F0의 역할은 매우 한정적으로 두 지역 모두 VOT만으로는 판단이 어려운 0 ms에서 높은 F0일 때 일부가 무성음으로 반응하는 정도였다. 이것은 2.에서 언급한 선행연구의 결과와도 일치한다. 도호쿠지방의 청자를 대상으로 한 유의미어에서는 F0가 0 ms보다 다소 넓은 범위(-16 ms에서 +8 ms 사이)에서 높은 F0를 무

성음, 낮은 F0를 유성음으로 판정하는 보다 적극적인 역할을 하는 것이 확인되었다. 다만 F0 효과는 VOT 효과에 비하면 영향력이 상대적으로 작아서 통계적으로 유의하기는 하나(VOT: $\beta = 0.180$, $p < 0.001$, F0: $\beta = 0.07$, $p < 0.001$) 아직은 VOT와 같은 정도의 카테고리 지각이 F0에서도 이루어지고 있다고는 보기 어려울 듯하다. VOT와 F0를 나누어서 살펴보자.

그림 6은 VOT와 F0별 응답률이다. 그림에 표시된 각각의 VOT값과 F0값에 대한 F0와 VOT는 그림 2, 그림 4, 그림 5에서 제시한 각 단계(10-11단계)를 모두 합한 것이다.

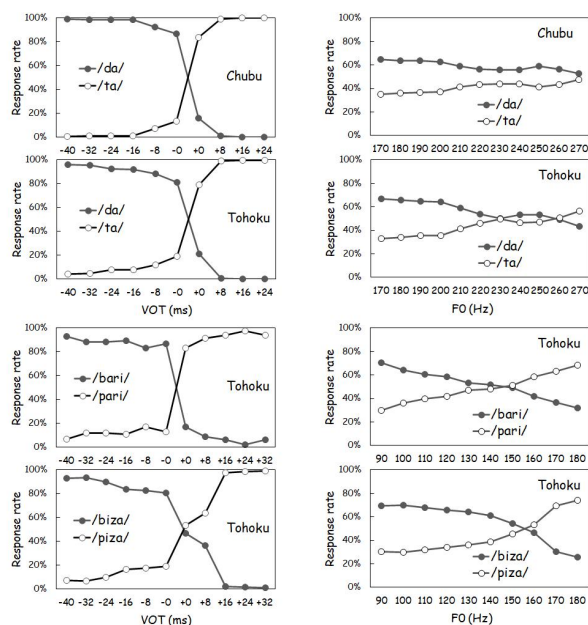
무의미어의 F0는 특히 특정 F0값(여기서는 220 Hz) 이상에서 유성음과 무성음의 응답률이 모두 50% 전후로 나타나 F0가 충분히 기능하고 있지 못하는 것을 알 수 있다. 반면, 유의미어는 특정 F0값(여기서는 150 Hz)을 경계로 응답률이 교차하여 높은 F0에는 무성음, 낮은 F0에는 유성음으로 정연하게 반응하고 있다. 다만 응답률 자체는 최대인 경우에도 74%에 머물러 VOT에서 보이는 100%의 예는 보이지 않는다. 다시 말해, 무의미어 /ta-da/는 주부지방, 도호쿠지방 모두 거의 VOT만으로 폐쇄음 지각이 이루어지나, 유의미어 /pari-bari/와 /piza-biza/는 VOT만으로는 부족한 경우가 있어 그 부족한 부분을 F0가 보완하고 있으며, F0 역할이 한정적이기는 하나 무의미어의 경우보다는 적극적으로 유성음, 무성음의 구별에 관여하고 있다고 할 수 있겠다.

이처럼 무의미어와 유의미어에 나타난 F0 역할의 차이는 통계적으로도 유의할 것으로 보이나 이미 언급한 것처럼 자극음의 F0의 범위가 다르기 때문에 처리가 불가하여 통계적으로는 확인이 안 된 상태이다.

5.2. 개인차와 혁신 청자(Innovative-Listener)

도호쿠지방의 화자가 다른 지역과는 달리 생성에서 어두 폐쇄음의 음향특성으로 VOT와 F0를 거의 같은 정도로 사용하고 있는 것을 고려하면(Byun, 2021a) 지각에서도 지각특성으로 VOT와 F0를 모두 사용할 듯하나 선행연구(Byun, 2021b; Gao et al., 2019)나 본고의 무의미어에서 확인된 것처럼 F0 사용은 VOT가 충분히 기능할 수 없는 경우에만 극히 한정적으로 사용되었다. 한편 유의미어에서는 대체로 양값의 VOT와 높은 F0의 자극음을 무성음, 음값의 VOT 또는 짧은 VOT(short-lag)와 낮은 F0의 자극음을 유성음으로 판단하는 경향을 보였는데 모든 청자들이 이런 식으로 인지하는 것은 아니다. 실험 결과를 개인별로 보면 폐쇄음 지각은 VOT만을 사용하는 청자, 정도의 차이는 있으나 VOT와 F0를 모두 사용하는 청자로 나뉜다. 그리고 그 중 F0를 적극적으로 사용하는 몇몇 청자의 결과가 유의미어 결과 전체에 영향을 끼친 것으로 보인다.

부록 1-4에 개인별 결과를 제시하였다. 도호쿠지방의 청자 01과 02는 무의미어 /ta-da/에서 VOT가 음값일 때 높은 F0에 대한 무성음 판정이 다른 청자들보다 두드러진다. 유의미어 /pari-bari/와 /piza-biza/에서는 청자 01과 02를 포함한 다른 청자들이 높은 F0뿐만 아니라 낮은 F0에도 반응한 것을 볼 수 있다. 반면에 같은 도호쿠지방의 청자라도 F0에 거의 반응하지 않는 이들도 있으며, 높은 F0가 아닌 낮은 F0에 보다 적극적으로 반



VOT, voice onset time; F0, fundamental frequency.

그림 6. VOT와 F0별 응답률
Figure 6. Response rates by VOT and F0

응하는 이들도 있다. 개인 내에서도 어휘에 따라 F0 사용에 차이가 있는 것은 폐쇄음 지각에 개인차 외에도 어휘차가 존재하는 것을 뒷받침한다. 다만 청자 01과 02는 어휘와 상관없이, 어휘의 의미 유무와 상관없이 F0 사용에 있어서 일관된 경향을 보여주고 있어 저자는 이들이 지각에 F0 사용을 획득한 혁신 청자(innovative listeners as an early adopter)가 아닌가 의심해 본다.

이미 언급한 것처럼 도호쿠지방에서 F0가 생성의 음향특성으로 중요한 역할을 하는 것은 젊은 세대로, F0 차이는 옛날 세대에서는 보이지 않는 새로운 특징이다(Takada, 2021). 젊은 세대에서 유성음, 무성음 사이의 F0 차이가 분명해져 청자가 인지할 수 있는 환경이 만들어지고 그것을 몇몇의 혁신 청자가 받아들여 인지하기 시작한 것이 아닌가 싶다. 이들을 중심으로 F0 역할이 강화되어 F0를 사용한 폐쇄음 지각이 언어공동체 전체로 퍼지게 된다면 VOT와 마찬가지로 F0가 지각특성으로 자리잡게 될 것이다. 또한 지각특성으로 인지가 가능해지면 생성에서도 의식적으로 사용할 수 있고 이로써 언어변화의 토대가 마련되게 될 것이다. 물론 이 시나리오가 실현되기까지는 상당한 시간이 걸리겠지만 현재 일본어에 나타나는 분절음의 F0에 관한 변화는 이 과정의 시작점에 있는 것이 아닌가 추측해 본다.

상기 시나리오는 지각특성으로 F0가 확립되는 데에 다른 저해 요인이 없는 것을 전제로 하지만 실제로는 고려해야 할 사항이 있다. 일본어는 한국어의 경상도 방언과 마찬가지로 성조(피치 악센트)를 갖는다. 어휘를 구별하는 초분절음의 F0 차이가 이미 존재하는 상태에서 분절음에 사용되는 F0 차이가 강화된다면 서로가 충돌을 일으킬 가능성이 있다. Gao & Arai(2019)는 발화에 나타나는 어두 폐쇄음의 F0 차이는 onset에만 한정되어 어휘 악센트의 F0 차이에는 영향을 주지 않는다고 하였으나 지각에서는 어떻게(본 연구에서는 HL음조만을 사용했기 때문에 성조의 영향 자체는 검토 대상이 아니었다). 분절음 지각에 사용되는 F0 차이가 초분절음 지각에 사용되는 F0 차이에 영향을 준다면 위에서 추측한 지각특성으로서의 F0 확립은 기대하기 어려울 것이다. 이에 대해서는 무악센트 지역을 포함한 다른 지역의 청자들을 대상으로 더 조사가 필요하겠다.

6. 결론

일본어 어두 폐쇄음의 지각특성으로서 VOT와 F0의 역할을 밝히기 위해 원음의 VOT와 F0를 재합성한 자극음을 사용하여 도호쿠지방 청자를 대상으로 조건을 달리한 지각실험을 실시하였다. 결과에서는 무의미어의 경우 지역차(도호쿠 지방 vs. 주부 지방)는 유의하지 않았으나 유의미어에서는 어휘에 따라 F0 사용에 유의한 차이가 있었으며 이러한 차이는 F0를 적극적으로 사용하는 몇몇의 청자들에게서 기인하는 것으로 밝혀졌다. 논의에서는 이들이 도호쿠지방의 혁신 청자들로 여겨지며 이들을 중심으로 폐쇄음 지각에 F0 사용이 일반화된다면 지각특성으로서 F0가 확립될 것이라고 추론해 보았다.

본 연구에서는 생성에서 파열 전 성대 진동을 동반하지 않고 유성음도 양값의 VOT를 갖는 도호쿠지방 청자를 대상으로 하

였으나 일본어 전체의 지각특성을 밝히기 위해서는 다른 지역의 청자들에서도 유의미어에서 F0가 지각특성으로 유효한지를 확인해 봐야 할 것이다. 더불어 지각에서 F0 사용에 어휘 차이가 있는 것은 본 연구에서 처음으로 확인된 사항으로 이러한 차이가 사용빈도와 같은 어휘 특성과 관련이 있는지에 대해서는 더 상세히 살펴봐야 할 것이다.

References

- Bang, H. Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T. J. (2015, August). The effect of word frequency on the timecourse of tonogenesis in Seoul Korean. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK.
- Bang, H. Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T. J. (2018). The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics*, 66, 120-144.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.
- Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer (version 6.2.22) [Computer program]. Retrieved from <http://www.praat.org/>
- Boersma, P., & Weenink, D. (2022). Praat: Doing phonetics by computer (version 6.2.23) [Computer program]. Retrieved from <http://www.praat.org/>
- Byun, H. G. (2018, June). Acoustic parameters of Japanese word-initial stops: VOT and post-stop F0. *A handout distributed in the 337th Regular Meeting of the Phonetic Society of Japan*. Kobe, Japan. Retrieved from <https://researchmap.jp/byun/presentations/18917628?lang=en>
- Byun, H. G. (2021a). Acoustic characteristics for Japanese stops in word-initial position: VOT and post-stop f0. *Onsei Kenkyū*, 25, 41-63.
- Byun, H. G. (2021b). Perception of Japanese word-initial stops by native listeners. *Phonetics and Speech Sciences*, 13(3), 53-64.
- 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.
- 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.
- R Core Team. (2022). R: A language and environment for statistical computing (version 4.2.2) [Computer software]. Vienna, Austria:

- R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- 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.
- Takada, M. (2004). The +VOT tendency in initial alveolar plosive/d/in Japanese and speakers' age. *Onsei Kenkyū*, 8(3), 57-66.
- Takada, M. (2011). Research on the word-initial stops of Japanese: Synchronic distribution and diachronic change in VOT. Tokyo, Japan: Kurosio.
- Takada, M. (2021). Generational differences in acoustic parameters for voicing contrast in the Tohoku dialect. *Bulletin of the Faculty of Letters of Aichi Gakuin University*, 50, 75-87.
- Tamura, S., Ito, K., Hirose, N., & Mori, S. (2018). Psychophysical boundary for categorization of voiced-voiceless stop consonants in native Japanese speakers. *Journal of Speech, Language, and Hearing Research*, 61(3), 789-796.

• 변희경 (Hi-Gyung Byun) 교신저자

국제교양대학 국제교양학부 준교수

Faculty of International Liberal Arts, Akita International University

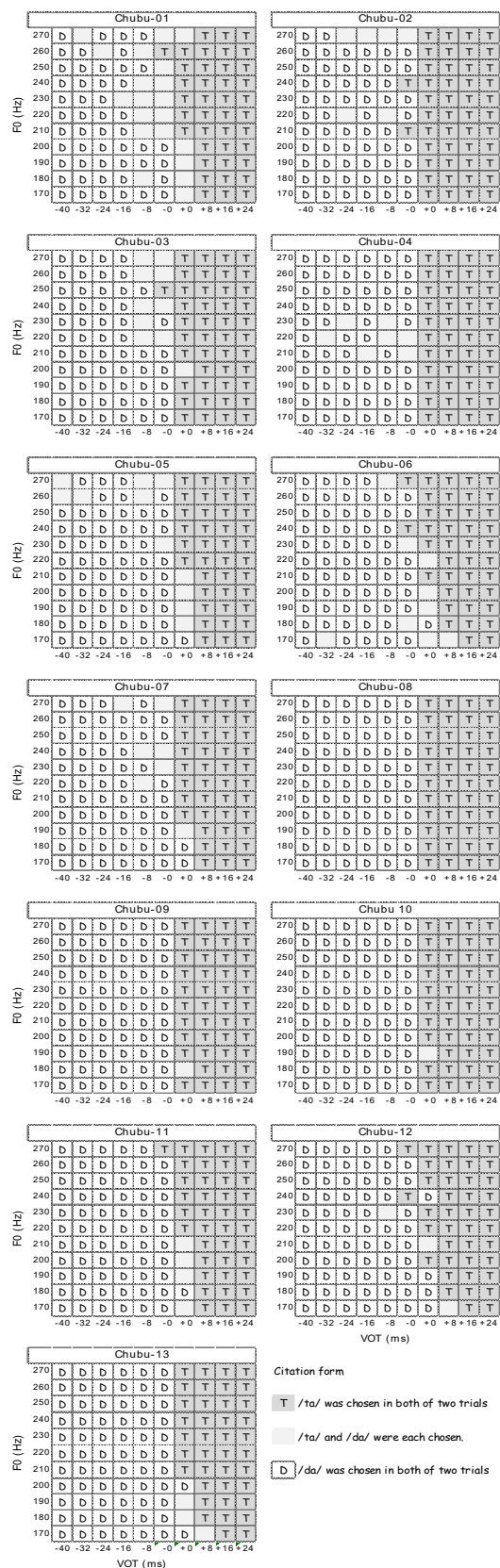
Yuwa, Akita-City, Akita 0101292, Japan

Tel: +81-18-886-5973

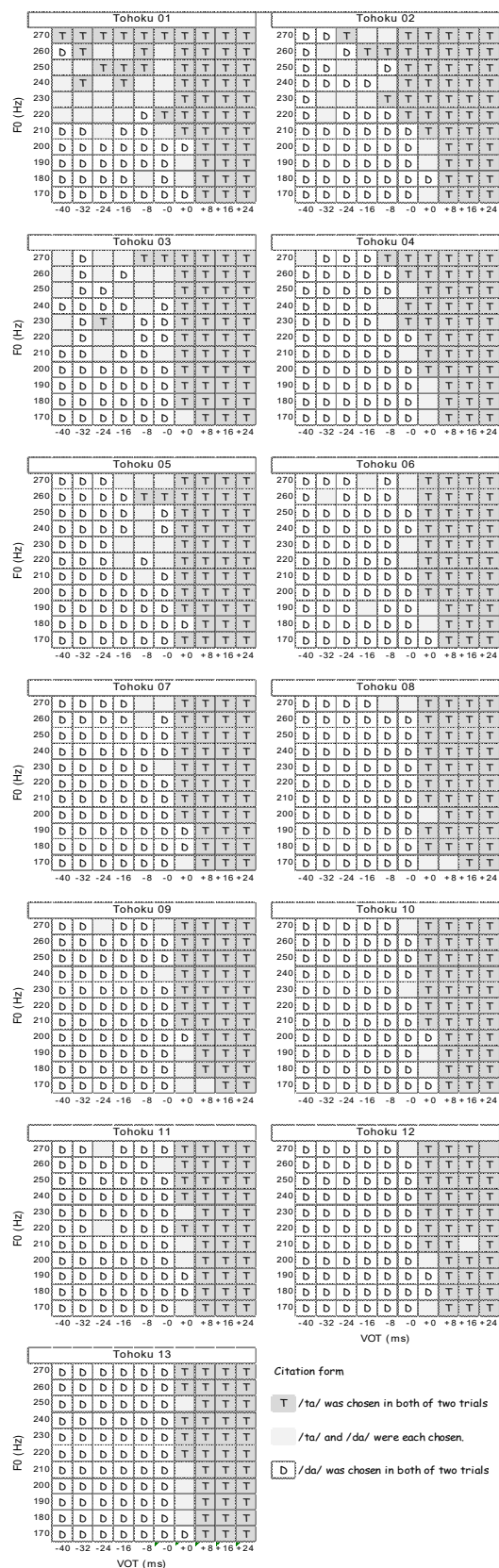
Email: byun@aiu.ac.jp

관심분야: 음성학, 한일음성 대조연구, 음성교육

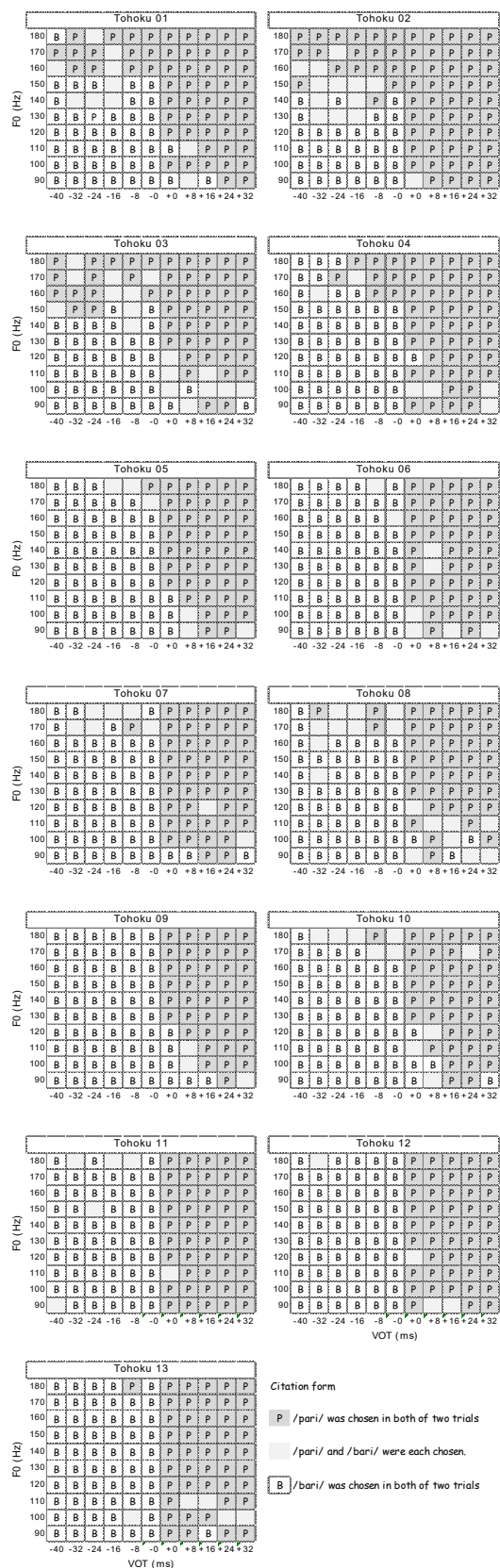
부록 1. /ta-da/ 주부 지방 개인별 결과



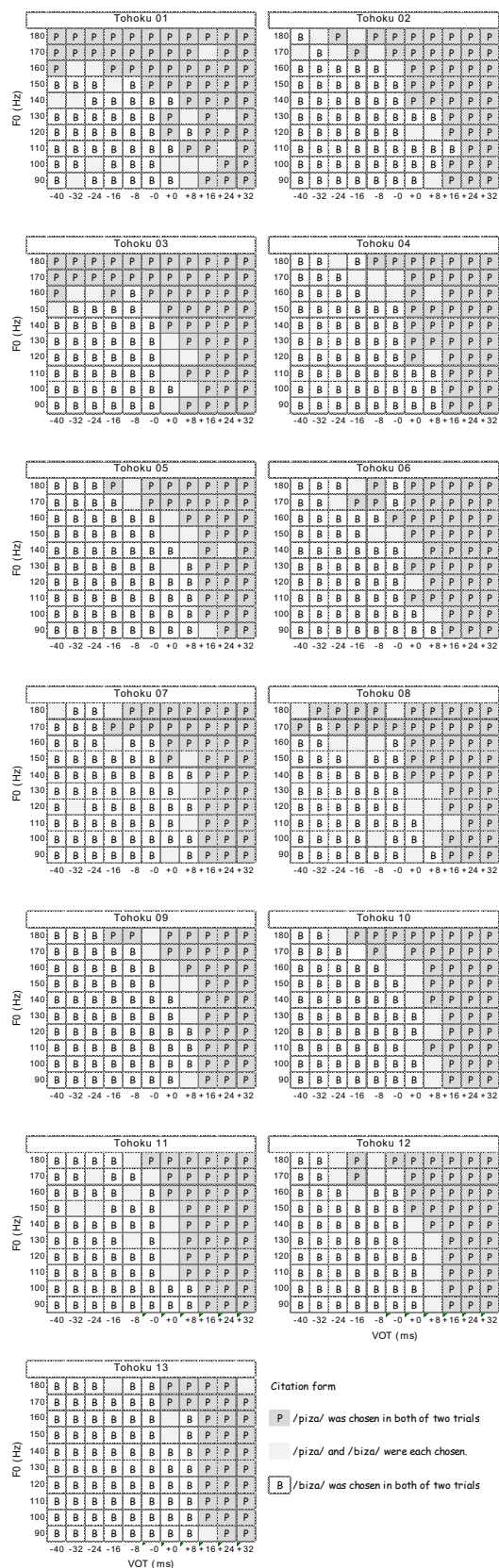
부록 2. /ta-da/ 도호쿠 지방 개인별 결과



부록 3. /pari-bari/ 도호쿠 지방 개인별 결과



부록 4. /piza-biza/ 도호쿠 지방 개인별 결과



도호쿠 일본어의 폐쇄음 지각에 있어서 voice onset time(VOT)과 후속모음 fundamental frequency(F0)의 역할*

변 희 경

국제교양대학 국제교양학부

국문초록

일본어의 전통적인 어두 폐쇄음은 파열 전에 성대 진동을 동반하는 유성음과 파열 후에 약간의 기음을 동반하는 무성음으로 이분된다. 한편 도호쿠지방의 유성음은 어느 세대나 파열 전에 성대 진동을 동반하지 않고 무성화한 유성음으로 실현되어 다른 지역과 대조를 이룬다. 무성화한 유성음은 voice onset time(VOT)이 양값으로 나타나고 그러면 기존의 무성음의 VOT와 충돌하게 되어 카테고리 구별에 영향을 미치게 된다. 이에 대해 도호쿠지방의 화자는 생성 시에 다른 지역과는 달리 폐쇄음 구별에 후속 모음의 fundamental frequency(F0)를 적극적으로 사용하는 것이 여러 연구에 의해 확인되었다. 본 연구는 인지면에서도 F0가 폐쇄음 구별에 중요한 역할을 하고 있는지를 밝히기 위해 VOT와 함께 검토한 것이다. VOT와 F0를 재합성한 자극음을 사용하여 도호쿠지방 청자를 대상으로 조건을 달리한 여러 개의 지각실험을 실시하였다. 결과에서는 무의미어의 경우 지역차(도호쿠 지방 vs. 주부 지방)는 유의하지 않았으나 유의미어에서는 어휘에 따라 F0 사용에 유의한 차이가 있었으며 이러한 차이는 F0를 적극적으로 사용하는 몇몇의 청자들에게서 기인하는 것으로 밝혀졌다. 논의에서는 이들이 혁신 청자들로 여겨지며 이들을 중심으로 폐쇄음 지각에 F0 역할이 일반화되고 지각특성으로서 F0가 확립될 가능성에 대해 추론해 보았다.

핵심어: 도호쿠 일본어, 어두 폐쇄음, 지각특성, voice onset time(VOT), 후속 모음 fundamental frequency(F0)

참고문헌

변희경(2021b). 모어청자에 의한 일본어 어두 폐쇄음의 지각. *말
소리와 음성과학*, 13(3), 53-64.

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