creative commons
COMMONS DEED

저작자표시-비영리-동일조건변경허락 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

동일조건변경허락. 귀하가 이 저작물을 개작, 변형 또는 가공했을 경우에는, 이 저작물과 동일한 이용허락조건하에서만 배포할 수 있습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

이학석사학위논문

의사결정 모델을 이용한 음소 분류과제의 선택과 반응시간 예측

Predicting Choices and Timings of Phoneme Categorization with a Perceptual Decision Model of Phonemic Processing

2013년 2월

서울대학교 대학원

뇌인지과학 전공

김 진 영

Predicting Choices and Timings of Phoneme Categorization with a Perceptual Decision Model of Phonemic Processing

지도교수 이 상 훈

이 논문을 이학석사 학위논문으로 제출함.

2012년 12월

서울대학교 대학원

뇌인지과학 전공

김 진 영

김진영의 이학석사 학위논문을 인준함.

2013년 2월

위 원 장 Randolph Blake (인)

부 위 원 장 이 상 훈 (인)

위 원 장 수 은 (인) S.E. Chang

Abstract

Predicting Choices and Timings of Phoneme Categorization with a Perceptual Decision Model of Phonemic Processing

JinYoung Kim
Brain and Cognitive Sciences
The Graduate School
Seoul National University

Despite crucial roles of pre-lexical units in speech perception, modeling efforts so far have been heavily focused on information processing at lexical or post-lexical stages, impeding the mechanistic investigation of speech perception. Given this dearth of frameworks for studying pre-lexical units, the current study proposes a system-level neural model for phoneme classification. A lynchpin idea behind the proposed model is that the brain represents phonemes as probabilistic quantities, likelihoods. With this idea, our model bridges three well-known canonical computations in the brain – sensory encoding, likelihood decoding and evidence accumulation - along a cascade hierarchy of neural processing towards generating inputs to a next stage of speech perception. At the initial stage, sensory neurons with different tuning curves for physical properties relevant to phoneme discrimination compute individual likelihoods for the presence of those properties. Phoneme neurons at the following stage compute likelihoods for specific phonemes by summing the outputs of those sensory encoding neurons with weighting curves tuned for their preferred phonemes. At the final stage, evidence-accumulation neurons compute and accumulate over time evidence to reach a discrete phoneme classification by integrating outputs of phoneme neurons in a task-optimal

manner over time. The accumulation-to-bound mechanism operating at this stage translates probabilistic information represented in the phoneme neurons' output into concrete choices at a certain time. This translation allowed us to test the empirical viability of our model by assessing its capability of predicting actual patterns of choice fractions and reaction times exhibited by human listeners engaging in phoneme classification under various listening conditions. Using a small number of parameters, the model predicted not only the static, categorical structure of phoneme classification as a function of physical stimulus property, but also the adaptation-induced, dynamic changes in classification on an identical stimulus. Furthermore, the model was flexible enough to cover the wide range of individual differences in phoneme classification behavior. With these behavioral constraints in conjunction with neural and computational constraints exercised in model construction, our model provides a framework for studying neural mechanisms underlying initial stages of speech processing by generating hypotheses and predictions that are testable by neurophysiological and behavioral experiments.

……

Keywords: phoneme categorization, perceptual decision, speech perception, sensory encoding, likelihood decoding, neural model

Student Number: 2010-24020

Contents

Figures

Introduction

Speech perception is effortless. Imagine yourself watching a world-cup final soccer match in a huge stadium packed with loud spectators. Your eyes are quite busy chasing a bouncing ball and top-notch athletes' spectacular movements around it, but your ears are not less. Despite constant ear-tearing waves of sounds coming from several thousand different individuals and instruments cheering for players on the ground, you somehow manage to listen to online radio streams of sport commentators' busy chats over the game through your left ear while reacting to occasional comments made by your friends sitting on your right. Our daily life experiences, like the one illustrated above, readily testify that the human brain must have implemented very sophisticated neuronal computations that would ultimately lead to seamless, online translations of acoustic input streams into meaningful linguistic entities with a great degree of robustness, precision and speed.

Among those neuronal computations for speech perception, one foremost crucial step is translating acoustic input streams into pre-lexical categories (Obleser & Eisner, 2009a). From a computational viewpoint, the forming of a limited number of abstract phonological representations at an early stage of speech processing can help address several fundamental problems in speech perception, including the 'invariance problem(Perkell & Klatt, 1986)'- a task of accomplishing perceptual constancy in a high degree of variability in speech sensory input (Kraljic, Brennan, & Samuel, 2008). In the example situation above, imagine that you have never heard the commentators before, and they speak English with a strong Scottish accent, to which you are unfamiliar. You can learn to recognize his speech much efficiently if the brain updates only a limited number of sub-word units based on phonological features unique to his pronunciation rather than if it has to adjust an entire set of word representations on a case-by-case basis.

This computational importance of pre-lexical categorization forced many

psycholinguistic or computational models of speech perception to adopt, either implicitly or explicitly, pre-lexical representations as primitive input to their lexical processing system (McClelland & Elman, 1986; Norris, McQueen, & Cutler, 2000). However, in majority, those models have been developed with a focus on the lexical processing stage without specifying mechanisms of pre-lexical processing. In this regard, there is even no clear consensus about the precise type of representations at the pre-lexical level (McQueen 2005 Handbook of cognition). This is not surprising given the meager, relative to lexical-level studies, amount of empirical or computational studies on spoken language processing at pre-lexical stages. In general, structural and functional properties of inputs can greatly constrain the way any given systems process those inputs to achieve their computational goals. Hence, the lack of mechanism-level understanding of pre-lexical representations imposes a fundamental limit on those models. The work presented here was motivated to advance the mechanism-level understanding of pre-lexical processing by predicting human observers' categorical responses to phoneme stimuli, which are widely believed to be one of the most likely candidates for pre-lexical categorization units (McClelland & Elman, 1986), with a model that is constrained both by computational optimality and by neural plausibility.

Our model is inspired by recent empirical observations and conceptual advancements in visual neuroscience about how sensory neurons encode stimuli in their population activity and downstream neurons decode task-directed information from that population activity, and how the outcomes at those sensory encoding and decoding stages translate into optimal decision behavior (see (Dayan & Abbott, 2001; Gold & Shadlen, 2007; Pouget, Dayan, & Zemel, 2003) for review). At the core of our model are two key conceptual frameworks that we borrowed from those visual neuroscience studies and adapted to explain pre-lexical categorization. First, we posit that pre-lexical information is represented in probabilistic values, not in unambiguous and deterministic values as often assumed by many pre-existing cognitive models of speech perception. To be specific, those probabilistic values are reflected in a set of high-tier decoding neurons that read out likelihoods for pre-lexical units from a given population activity of

upstream neurons encoding acoustic features of speech signals (hereafter, those encoding-stage and decoding-stage neurons will be referred to as 'acoustic feature (AF)' and 'phoneme likelihood (PL)' neurons, respectively). Second, to convert those PL neurons' responses into categorical choices made by observers with a certain temporal lag, our model has a separate decision unit. A decision neuron (hereafter referred to as an 'evidence accumulation (EA)' neuron) extracts decision evidence from PL neurons' output in a task-dependent manner and accumulates over time the evidence until it hits a bound, which terminates the decision process and triggers motor execution (Gold & Shadlen, 2007). We adopted 'EA' neurons, which are prevalent throughout the brain and thus considered as one of the canonical neural computations(Carandini, 2012), because the task-optimal nature of their evidence abstraction (Jazayeri & Movshon, 2006) and its capability of resolving speed-accuracy tradeoff – making a most accurate choice for a given speed or a fastest choice for a given accuracy - (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006) together address the issue of optimality in pre-lexical categorization behavior. Our model will be referred to as the 'likelihood model of phoneme classification', likelihood model in abbreviation.

To verify the viability of the likelihood model, we examined how capable it is of capturing dynamical changes in human observers' phoneme categorization responses to artificial phoneme stimuli under various hearing conditions. In the first experiment, we asked observers to classify stimuli into one of three stop-consonant phonemes ('/da/', '/ba/', and '/ga/') while systematically varying acoustic features of the stimuli along a path in a multidimensional spectral space known to define the perceptual spectrum encompassing those three target phonemes. This allows us to evaluate the robustness of the model in mapping acoustic stimuli onto phoneme categories in a manner reflecting '*steady-state*' phoneme representations idiosyncratic to individual observers. In the following experiment, we challenged the model further by testing whether it can predict '*dynamic-state*' phoneme representations subject to temporal contextual modulation. To perturb observers' intrinsic representations of phonemes temporarily, we exploited adaptation effects, which have been used as a powerful behavioral tool to infer

mechanisms of neural coding of visual stimuli (Lee & Lee, 2012; Schwartz, Hsu, & Dayan, 2007). With this adaptation protocol, we monitored the changes in categorization of physically identical test stimuli while varying adapting stimuli.

Then, we fit the likelihood model to the categorization behavior of individual observers in the two experiments in terms of two major types of decision metrics, choice fraction and reaction time (RT). Using a fairly small number of parameters with biologically plausible ranges of values, the model successfully generated the phoneme likelihoods, for both static (Exp 1) and dynamic (Exp 2) state representations of given phoneme stimuli, in the activity of PL neurons, which were translated by EA neurons into choice fractions and RTs matched to the observed ones, respectively. In addition, in-depth inspections and simulations of the model, which was designed to reflect specific neuronal populations' functional and computational properties, revealed two important aspects of phoneme categorization. First, between its hierarchically organized neuronal components, one encoding acoustic features and the other decoding phoneme likelihoods, the model decisively indicated the former as an origin of adaptation effects, a conclusion dovetailed with that of adaptation studies on visual motion (Kohn & Movshon, 2003; Lee & Lee, 2012). Second, thanks to its decision-stage model component implementing 'accumulation-to-bound' computation, the model could provide a mechanistic account for intriguing patterns of RT variability across trials and across individuals, which have been either neglected or left unexplained by previous studies, based on optimal decision under the speed-accuracy trade-off context.

Likelihood model of Phoneme Classification

We assumed that, in the listener's brain engaging in speech perception, phonemes are represented as probabilistic quantities that are constructed through a neural process consisting of sensory encoding and phoneme likelihood decoding (Fig 1). Given a sound input, θ_s, made by the speaker with intention of generating a *true* phoneme, ψ_s, a population of *sensory neurons* (SNs) encodes θ_s in their responses, $sr_{i=[1,n]}(\theta_s)$. If we assume that each of those sensory neurons is broadly tuned as a function of θ, $sr_i=f_i(\theta)$, it is well established from previous work [Seung, H.S. & Sompolinsky, H 1993; Jazayeri & Movshon 2006)] that individual neuron's responses sr in a given trial and their tuning functions can jointly represent the likelihoods of the stimulus θ_s in the log space:

$$bg\ L_i(\theta_s) = sr_i bg\ f_i(\theta_s) - f_i(\theta_s) - \log(sr_i!) \qquad (1)$$

Thus, at the initial stage of our model, the sound θ_s evokes population responses of the SNs tuned to different values of θ, sr, which in turn collectively encode θ_s by gauging how likely θ_s is present in an incoming physical input at a given moment.

In the second stage, there exists a set of *phoneme neurons* (PNs) that decode the likelihoods of particular phonemes from the population responses of the SNs at the previous stage. The computation achieved by these PNs is a reverse engineering of the speech production process, in which the speaker generates a sound stimulus θ_s with an intention of delivering a phoneme ψ_s to the listener. The model assumed that each of the phoneme neurons performs this reverse engineering by computing the likelihood of its preferred phoneme ψ_s, which can be formalized as a weighted sum of the SN population's likelihood representation of θ_s:

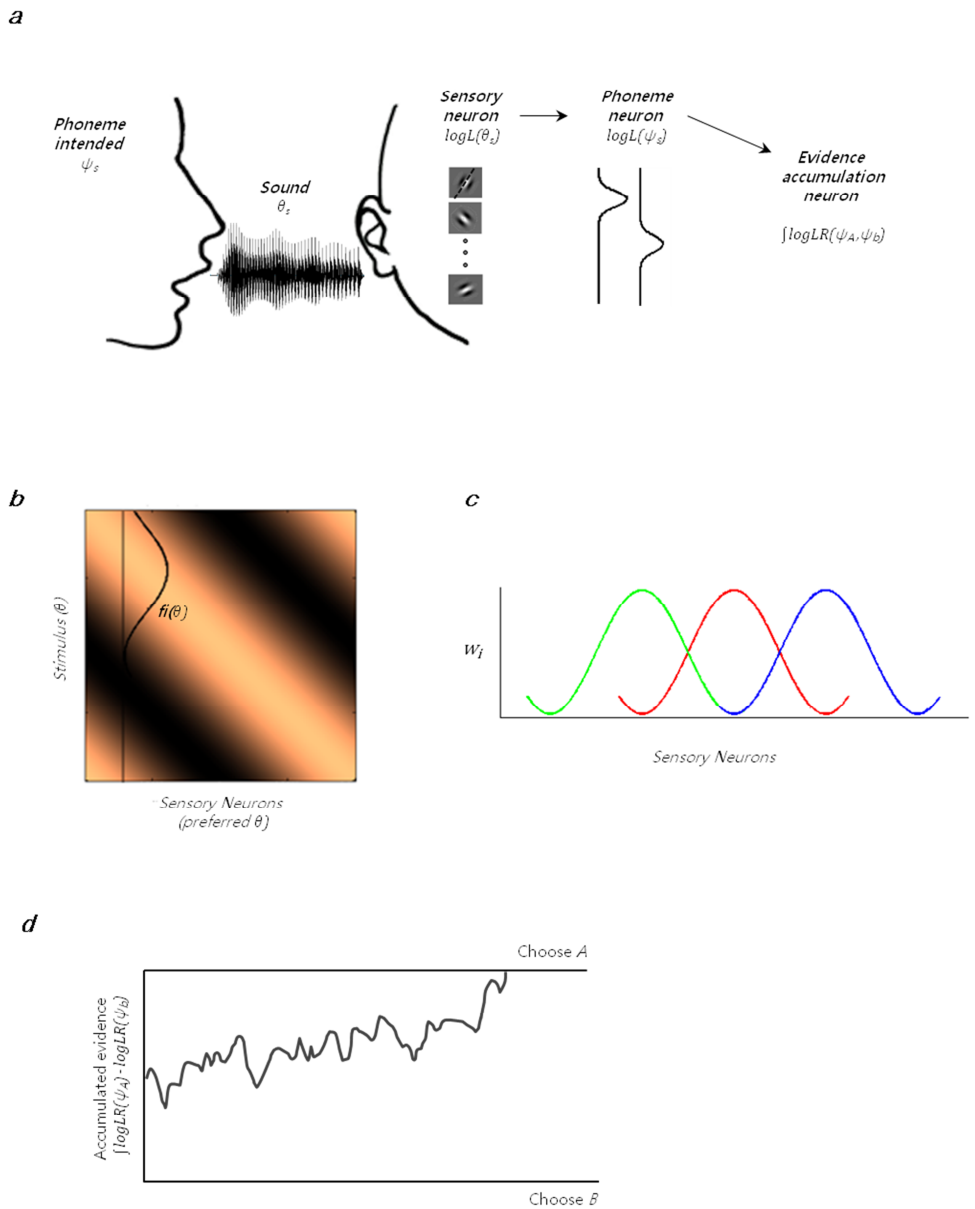


Figure 1. Neural model for phoneme perception. (a)Neural process of phonemic decision making. (b)A receptive field of sensory neurons (c)Likelihoods of three phonemes .(d) Evidence accumulating neurons.

$$\log L\ (\psi_s) = PN(\theta_s) = \sum_{i=1}^{N} w_i \log\ L_i(\theta_s)$$

$$= \sum_{i=1}^{N} w_i sr_i \log\ f_i(\theta_s) \qquad (2)$$

From a previous work (Jazayeri & Movshon 2006), we know that the last two terms of equation (1) can be ignored here because the sum of population tuning functions and responses are both independent of θ. Here, a weighting function w is a set of pooling weights assigned to the individual SNs. Conceptually, w can be understood as a neural implementation of a reverse of the probability distribution of sounds θ when she or he intend to speak a given phoneme ψ_s. In other words, individual SNs' contributions to the likelihood of a given phoneme are determined by their tuning function f_i. There are PN clusters with different phoneme preferences, and the number of those clusters matches the limited number of phonemes for a spoken language used by the speaker and listener.

Although the second stage, where the likelihoods of phonemes are formed, comprises a core part of the model, phoneme processing requires one more step to complete its ultimate goal in the context of speech perception, categorization of sensory inputs into discrete pre-lexical units. The model implemented the process of categorical perceptual judgment by adopting evidence-accumulating neurons (EANs) [Yang & Shadlen, 2007; Gold & Shadlen, 2006], which perform two key computations. First, the EANs extract evidence for performing an impending task, e.g., identification or discrimination. The phoneme categorization task can be conceptualized as choosing one out of multiple known alternatives, phonemes. In the simplest version of categorization, where two phoneme are discriminated, the EANs can extract the task-optimal evidence from the PNs' outputs at the previous stage by computing the ratio of the two likelihoods, respectively represented by the two PNs with preferences for the two candidate phonemes. Because the likelihood ratio is equivalent to the difference of log likelihoods, the task-optimal evidence, E_T, can be extracted if an EAN combines the outputs of the two PN clusters with opposite signs:

$$E_T = \log LR\ (\psi_A, \psi_B) = \log L\ (\psi_A) - \log L\ (\psi_B)$$

$$= PN_A(\theta_s) - PN_B(\theta_s) \quad (3)$$

The second important computation is to accumulate E_Ts over time until the integrated evidence reaches one of the two bounds, assigned to the two candidate phonemes, respectively. In the formalism of accumulation to bound, the identity of a resulting pre-lexical unit is determined by which bound is hit by the integrated evidence whereas the timing of that pre-lexical decision is determined by when the bound is hit.

In summary, the phoneme likelihood model translates a given sound input into a categorical decision that is specified both in identity and in timing. While doing so as described in the above, the model achieves two important kinds of computational optimality, which relate to optimal behaviors demonstrated by human observers. First, the computation of E_T can be described as *optimal* in that it endows the model with the capability of extracting evidence for pre-lexical decision in a task-optimal manner (Jazayeri & Movshon, 2006). Second, the accumulation-to-bound computation helps the model efficiently achieves adaptive compromises between speed and accuracy of task performance, for example, by adjusting locations of decision bounds (Reddi and Carpenter, 2001). In addition, the phoneme likelihood model achieves neural plausibility by implementing those optimal computations based on several canonical neural computations that have been supported empirically in visual neuroscience (Carandini, 2011). Given this theoretically healthy set of features, the likelihood model was put to empirical tests on its ability to predict human subjects' phoneme categorization behavior in two different listening conditions.

Phoneme classification on a cyclic spectrum of stimuli varying in frequency modulation

We conducted the first experiment to examine whether the model is capable of describing a hallmark feature of phoneme perception, categorical perception. We asked listeners to classify stimuli into one of three stop-consonant phonemes (‘/da/’, ‘/ba/’, and ‘/ga/’) while systematically varying acoustic features of the stimuli along a path in a multidimensional spectral space known to define the perceptual spectrum encompassing those three target phonemes. Listeners’ phoneme classification responses to a given stimulus were assessed by two metrics: choice fraction (CF) and response time (RT). Then we evaluated the robustness of the model in mapping acoustic stimuli onto these two metrics in a manner reflecting ‘*steady-state*’ phoneme representations idiosyncratic to individual observers.

Methods

Subjects and apparatus

Fifteen Korean native speakers and one Korean-English bi-lingual speaker with normal hearing (four females; 18-28 years old) participated in experiments after providing written informed consent. Before conducting a main experiment, listeners participated in a hearing test, in which they performed a sound localization task (binaural discrimination) on pure-tone audio stimuli matched to experimental stimuli with frequency of 500 Hz ~ 4,000 Hz and amplitude of 25 dB stimuli that were generated. Only those who showed 100% performance on this screening test participated in the main experiments. Experiments were designed with E-Prime software in i-Mac and conducted in a dark quiet room. Auditory stimuli were presented through earphones (Etymotic Research ER-4B), with instructions being displayed on a monitor (HP

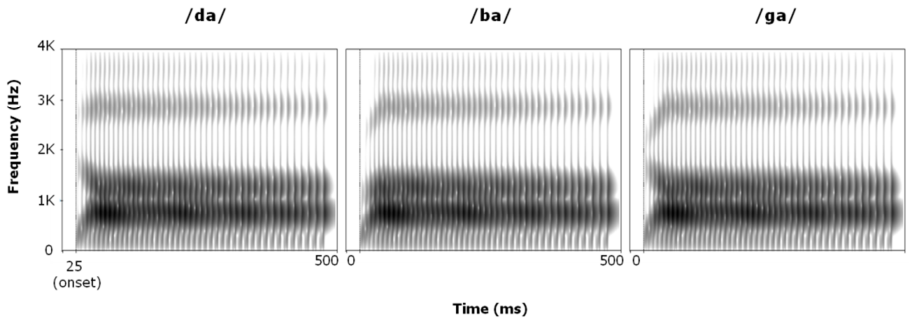
LP2065). Listeners' responses were recorded with a numeric keypad (SAMSUNG SNK2000).

Stimuli

Each stimulus, which was played for 475ms, consisted of six formants (Fig 2a). The fundamental frequency (F0) was 132 Hz at onset time (25ms), and then it fell to 120 Hz in 40ms. The starting frequency of the first harmonics (F1) was 200 Hz and reached its steady state 720 Hz at 50ms. The steady state values of the second (F2) and third (F3) were 1240 and 2850 Hz, respectively. The frequency values of the fourth, fifth, and sixth formants were maintained at 3650, 4500, and 4900 Hz respectively. Using PRATT software, we synthesized a cyclic spectrum of voiced stop consonant-vowel syllables by gradually varying the starting frequencies of the F2 and F3 components (Steinschneider et al., 1995), producing a total of eighteen syllables whose sounds smoothly change from '/da/' to '/ba/' to '/ga/' and then go back to '/da/'.

We defined a scale for those 18 stimuli in the following steps. First, the velocity of frequency modulation (VFM) at the initial 50-ms modulation period was calculated separately for the F2 and F3 components, resulting in two vectors of VFM. Second, we normalized those two vectors by transforming into z values. Third, we mapped the eighteen stimuli onto a space with axes defined by the two principal components, which were identified by applying the principal component analysis to the two normalized vectors of VFM values (Fig 2b). Finally, to make the scale reflect the physical similarity in that principal space (so that it can function as an interval scale), we calculated an angle of a vector connecting the space origin (0,0) and a given stimulus' coordinate (the colored lines in Fig 2b). This scaling procedure allows us to define the sound stimuli in a one-dimensional cyclic space by assigning angular values, θs. Hereafter, the physical property of experimental stimuli will be represented in this angular metric.

a



b

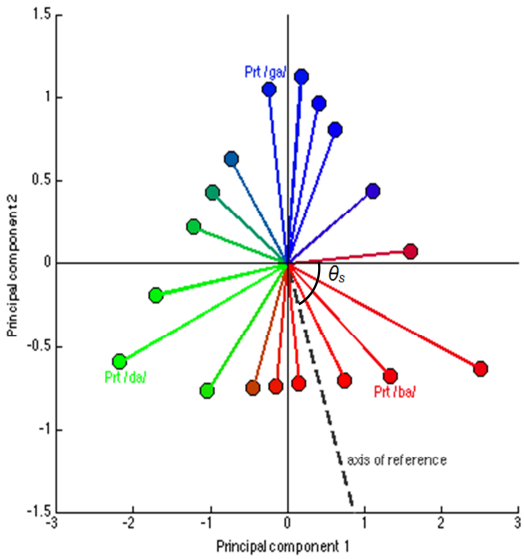


Figure 2. Synthesized stimuli for experiment 1. (a) Spectrograms of prototype sounds (/da/, /ba/, /ga/). (b) Eighteen stimuli in a space defined by Principle Component Analysis (PCA) of F2 and F3 velocity. Red refers to /ba/, green to /da/, and blue to /ga/.

Procedure

Before the main experiment, listeners heard three prototypical syllables (/ba/, /da/, and /ga/) repeatedly until they became familiar to synthesized stimuli. The prototypical syllables were identified by a pilot test as those leading to the highest fraction of choice for each phoneme category. Listeners proceeded to the main experiment when they correctly discriminated those prototypical stimuli in more than 27 out of 30 trials (>90%).

On each trial of the main experiment, listeners heard a single syllable stimulus and performed a two-alternative forced choice task (2AFC) by classifying it into one of two alternative categories. There were three type of trial blocks, differing only in terms of which pair of phoneme categories was used as choice alternatives: /da/ vs /ba/, /ba/ vs /ga/, and /ga/ vs /da/. In a given type of block, we presented only 6 neighboring stimuli that bridge between the two prototypical phoneme stimuli corresponding to alternative categories (e.g., θ = [-0.76, -1.36, 1.38, 1.04, 0.64,] for the /da/-vs-/ba/ block). Each listener repeated each block type twice, completing 6 blocks in total, which took 30 mins approximately. To have control over speed-accuracy trade off, listeners were instructed to complete their decision within 2,500 ms after stimulus onset. Listeners made responses by pressing a button on a keypad using one of the three right-hand fingers (index for /ba/, middle for /da/, and ring for /ga/).

Results

For the majority of listeners, categorical perception was evident both in CF and RT data. When plotted as a function of θ, the FC data exhibited two signature features indicating the presence of categorical classification. First, the CF curves had plateaus around prototype stimuli, more conspicuous in the /ba/ and /ga/ curves (red and blue lines, respectively, in Fig 3a) than in the /da/ curve (green lines in Fig 3a). Second, the CF

curves changed abruptly at around the boundaries between two given phoneme categories, with steep slopes, which were again more prominent in the /ba/ and /ga/ curves than in the /da/ curve. These two features indicative of categorical classification could be appreciated in the CF curves averaged across listeners, although the plateaus and slopes were narrowed and flatter due to data smoothing associated with averaging (Fig 3b).

The RT data almost mirrored the FC data (Fig 3c,d). The RT was fastest around the centers of the plateaus in the FC data and steeply increased with approaching the categorical bounds. We note that, although not observed for all of the listeners (probably due to the small numbers of trials), the RT tended to keep increasing even beyond the boundaries (the data points indicated by the arrows in Fig 3d). This monotonic increase in RT with decreasing prototypicality has not been explicitly demonstrated in previous studies.

To quantify these apparent anti-correlations between the CF and the RT data, we sorted the population-averaged CF and RT data as a function of stimulus prototypicality, that is a distance from the prototype stimulus ($\Delta\theta$) (Fig 3e,f). The correlation analysis confirmed the three main qualitative observations described above by resulting in the corresponding significant correlations: the negative correlation between the CF and the $\Delta\theta$ ($r=-0.79$, $p < 0.01$), the positive correlation between the RT and the $\Delta\theta$ ($r=0.82$, $p < 0.01$), and negative correlation between the CF and the RT ($r=-0.93$, $p < 0.01$).

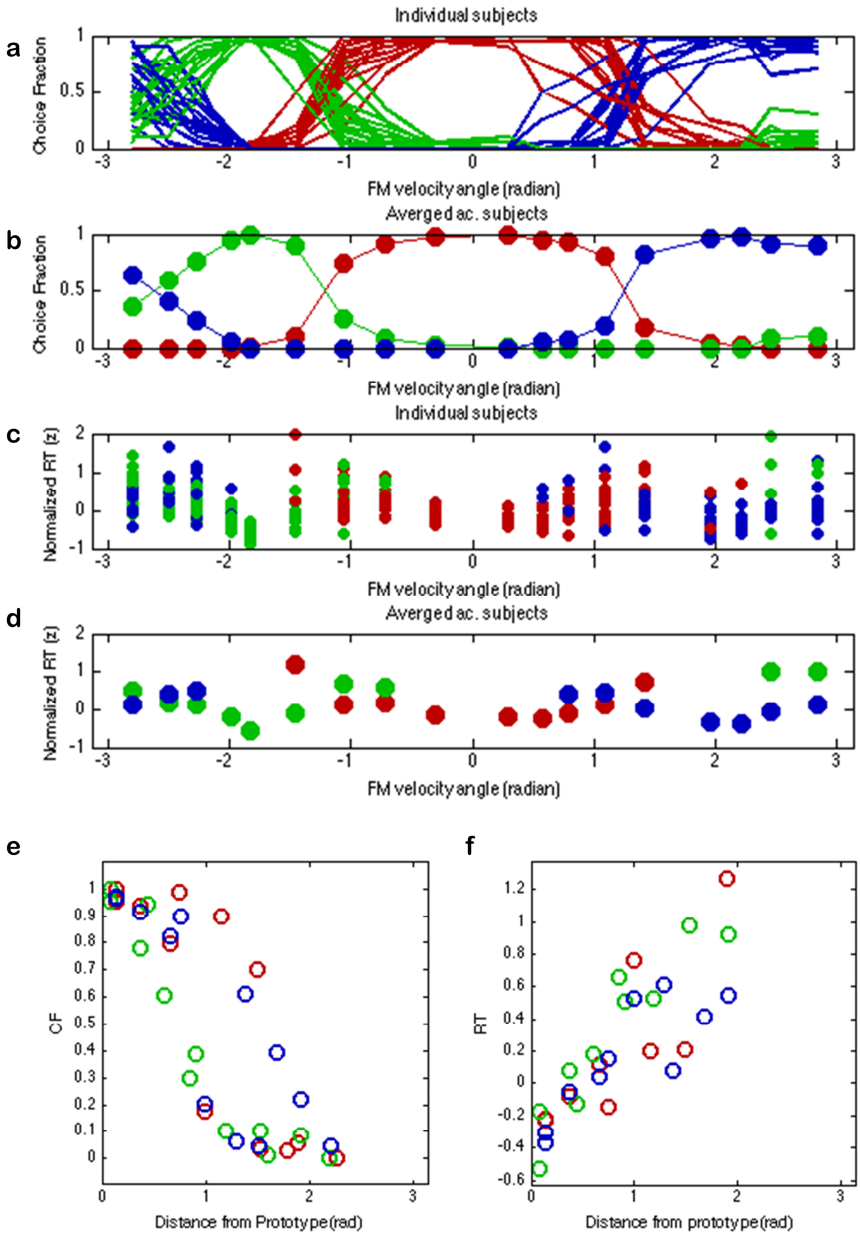


Figure 3. Behavioral result of experiment 1. (a) Identification function with proportion of choices. (b) Identification functions for individuals. (c) Mean normalized Response Time(RT) of each stimulus choice. (d) Mean normalized RT for individuals. Green stands for /da/, red for /ba/, and blue for /ga/. (e) correlation of choice fraction and $\Delta\theta$ (f) Correlation of Response time and $\Delta\theta$.

Dynamic changes in phoneme representation following adaptation

The second experiment was designed to evaluate how capable the proposed model is of capturing phoneme classification responses when listeners' intrinsic representations of phonemes were temporarily altered by temporal contextual modulation. To perturb listeners' intrinsic representations of phonemes in a laboratory, we modulated the temporal context of sound stimuli by exploiting an adaptation paradigm, which have been used as a powerful behavioral tool to infer mechanisms of neural coding of visual stimuli (Lee & Lee, 2012; Schwartz et al., 2007). With this adaptation paradigm, we monitored changes in phoneme categorization of physically identical test stimuli while varying adapting stimuli systematically.

Methods

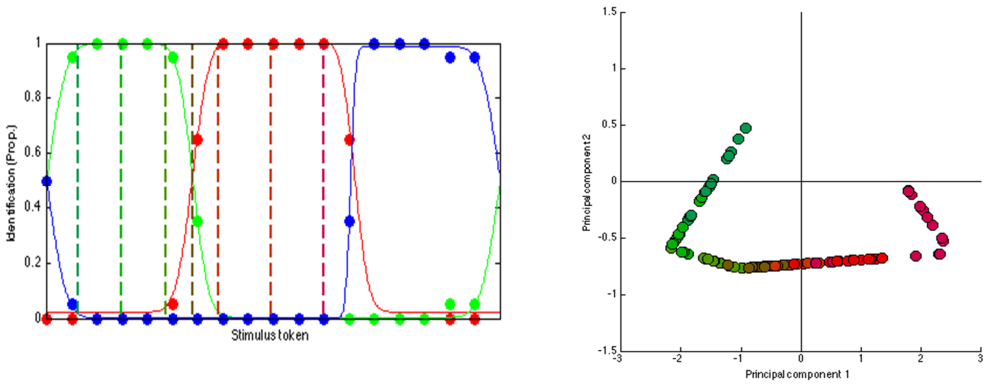
Stimuli

Seven stimuli were synthesized individually based on experiment 1 results. Three stimuli per category were synthesized to observe adaptation effect depending on different places on the category plateau, and an ambiguous sound between /ba/ and /da/ was generated as a test stimulus.

We made a model to fit choice performances from experiment 1 at first, and then found appropriate stimuli for experiment two with own specified criteria (fig 4A). A function below employed to define a place of each phoneme prototype.

$$ae^{\{b\times\cos(x-p)\}} + m = y$$

a ***b***



c

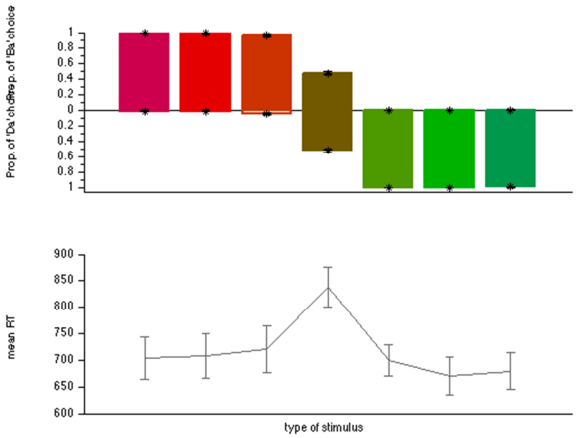


Figure 4. Stimuli for adaptation experiment. (a) Psychometric identification function of one subject and seven selected stimuli. (b) The location of selected stimuli in PCA space of F2 and F3. (c) Results of a pre-identification test with selected stimuli.

x refers to stimulus position in a polar space and y does to proportion of a certain-category choice. a, b, and m are free parameters. The location of prototype, p, was obtained by fitting data to the function, and it used as a prototype adaptor.

Afterward, we divided each category performance data into two parts with the prototype as the center and then fitted each part to the cumulative normal distribute function. Two 90% points on the obtained category were connected with a straight line, and then 1/10 and 9/10 points on the line were selected as positions of a proximal and a distal adaptor. A tester between /ba/ and /da/ was selected from a contact point of two categories. Furthermore, we did verification tests with stimuli near the obtained position, and selected appropriate stimuli that passed criterion (adaptors; above 95% choice response, ambiguous stimulus; the closest 50% response). A name of the adaptor (proximal or distal) was determined by relative position with the ambiguous stimulus.

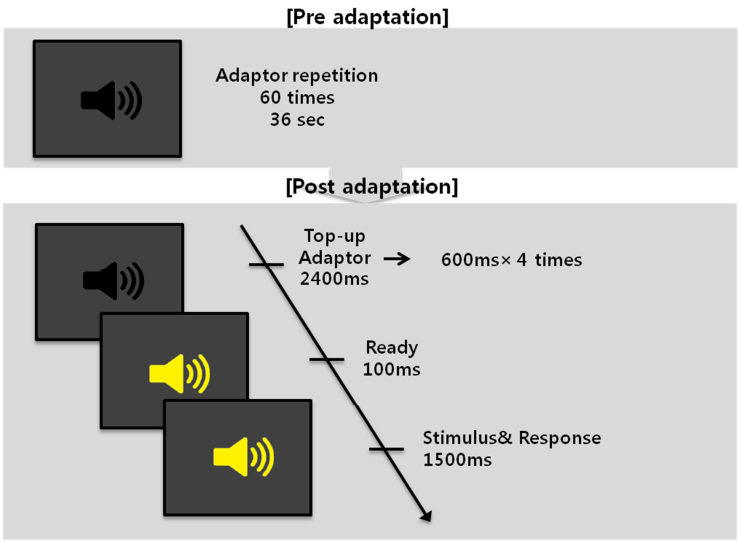


Figure 5. Procedure of an adaptation test. A session started with pre-adaptation that an adaptor repetitively played for 36 sec. Post adaptation followed starting with 4 repetitions of top-up adaptor and then a test stimulus presented. Subjects were instructed to response only when the speaker sign was bright (tester was presented)

Procedure

Nine blocks were provided in experiment 2 .One block was a simple classifying test (‘ba’ versus ‘da’) and the others were adaptation tests. In the classifying test, three test stimuli (prototype (prt) ba, da, and ambiguous(amb) stimulus) were randomly presented for 100 trials (the ratio of prt ba, prt da, and amb stimulus; 1:1:2). A design of adaptation tests was similar to those commonly used in vision research(Obleser & Eisner, 2009b). An adaptor was repeated 60 times for 36 s at the beginning of each block and 4 times for 2.4 s before each trial. The length of sound stimulus was 475 ms and that of interval was 125ms between repetitions. A tester provided in 100ms after the 4 repetitions of top up adaptor and this block was replayed for fifty times (Fig 5). Prt /ba/, prt /da/, and amb stimulus used as testers in adaptation tests (ratio; prt/ba/:prt/da/:amb = 1:1:8). Participants were instructed to identify the test sound appeared with a visual sign, but just listen carefully to adaptors without any response. In addition, they asked to respond the test sound within 1500 ms. Eight subjects were conducted adaptation task with four adaptors (proximal /ba/, distal /ba/, proximal /da/, and distal /da/) for two days, and the rest eight subjects were with six adaptors (prt/ba/ and prt/da/ added) for three days. One session took nearly 50 min.

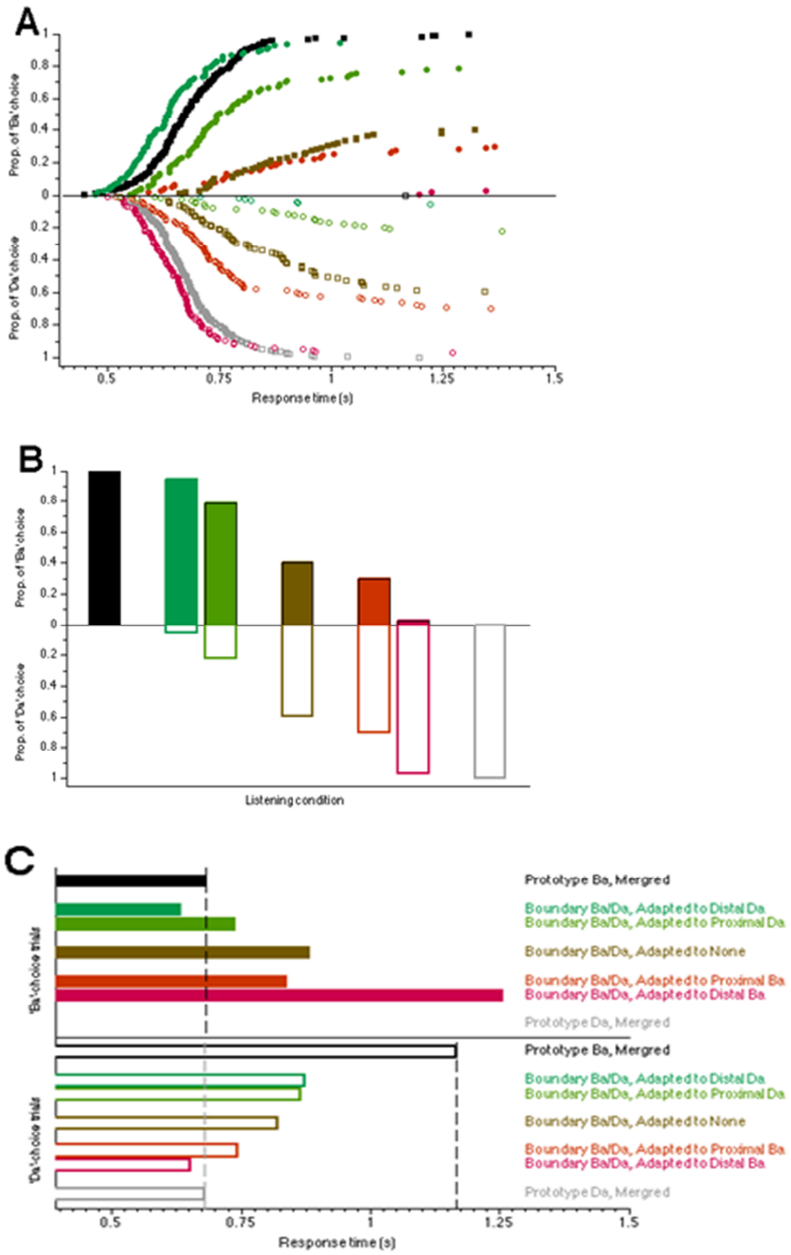


Figure 6. Behavioral results of experiment 2 with 4 adaptors from one particular subject. (a) RT distribution to proportion of choice. (b) Choice ratio for 7 different condition. (c) Mean normalized response time for choices according to adaptors.

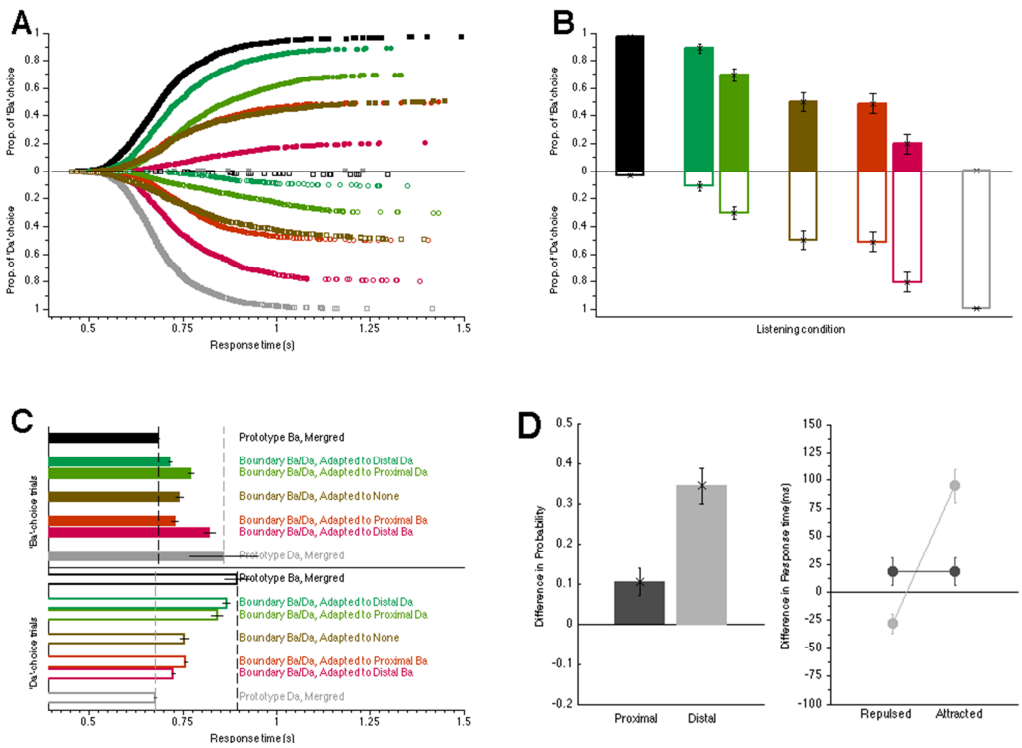


Figure 7. Behavioral results of experiment 2 with 4 adaptors from 8 subject. (a) RT distribution to proportion of choice. (b) Choice ratio for 7 different condition. (c) Mean normalized response time for choices according to adaptors. (d)Statistics for differences between proximal and distal condition.

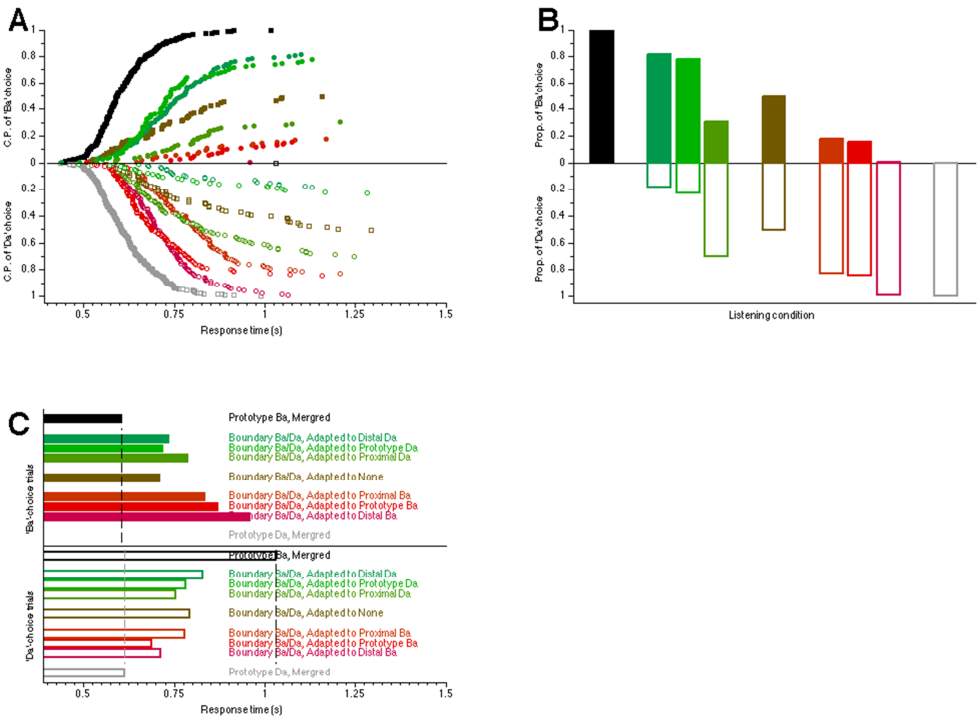


Figure 8. Behavioral results of experiment 2 with 6 adaptors from one particular subject. (a) RT distribution to proportion of choice. (b) Choice ratio for 7 different condition. (c) Mean normalized response time for choices according to adaptors.

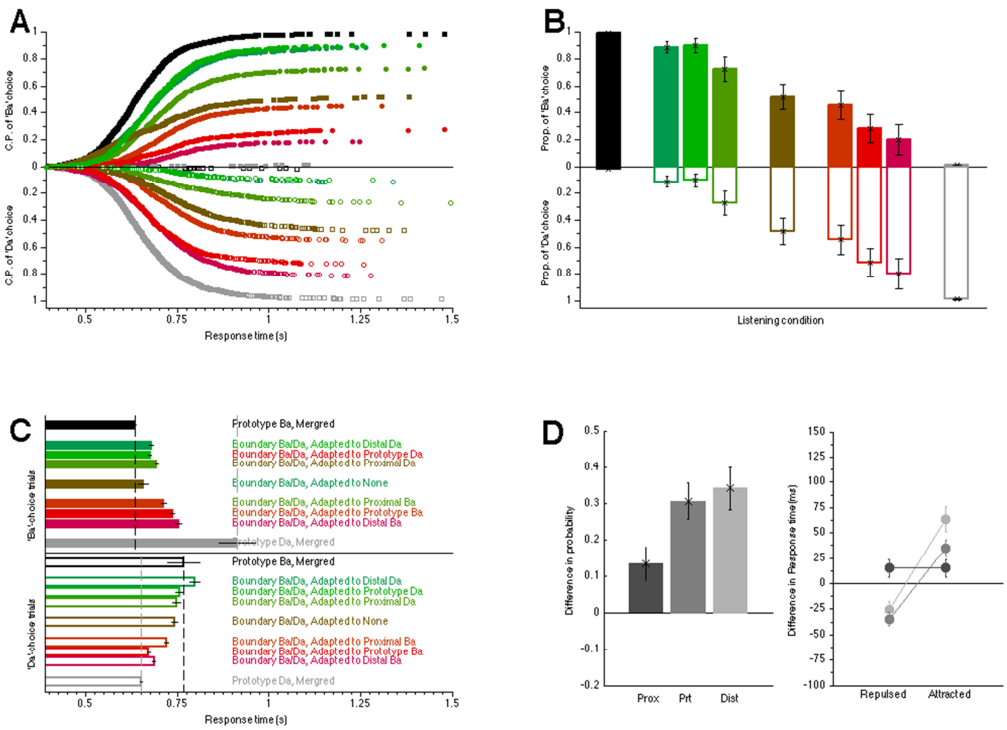


Figure 9. Behavioral results of experiment 2 with 6 adaptors from eight subjects. (a) RT distribution to proportion of choice. (b) Choice ratio for 7 different condition. (c) Mean normalized response time for choices according to adaptors. (d)Statistics for differences between proximal and distal condition.

Results

Seven stimuli (proximal /ba/ and /da/, distal /ba/ and /da/, prototype /ba/ and /da/, ambiguous /b?d?a/) for adaptation experiments were selected individually according to individual identification function from experiment 1 (fig 4a). To confirm the selected stimuli, pre-identification task was conducted before the experiment and the performance result was as figure4c we expected.

Eight subject were tested with four adaptors (proximal /ba/ and /da/, distal /ba/ and /da/) and three testers (prototype /ba/ and /da/, ambiguous /b?d?a/). An ambiguous tester (A-tester) placed between 'ba' and 'da' was affected by adaptation; subjects repetitively heard 'ba' before doing identification task with it, then it was more perceived as 'da' sound, while they heard 'da' before, then it was as 'ba' sound. Two tester, prototype 'ba' and 'da', on the other hand, were hardly influenced by adaptation. The adaptation effect to A-tester was, however, different according to types of adaptors. The adaptation effect was getting higher, when the adaptor was farther from the A-tester. That is, A-testers were more affected by distal adaptors than by proximal adaptors even though they perceived same sound in a pre-identification test. In addition, normalized mean response time decreased as adaptation effect increased (Fig 6,7).

We found that the adaptation effect was different in adaptors relative position to A-tester, despite adapters in the same category with same identification performance(Fig 8,9) . The result showed the possibility that the adaptation effect comes from a low level like acoustic channel not a phoneme level. It is also possible, however, that the proximal stimulus more sounds like a certain phoneme category than distal stimulus. The other eight subjects took the adaptation test with 2 more adaptors (prototype /ba/ and /da/) to examine this possibility.

The effect of adaptation from /ba/ category increased as an adaptor was far from the A-tester; distal /ba/ gave the highest effect, and prototype /ba/, proximal /ba/ in order. In the case of adaptation with /da/ category sounds, there was a significant difference

between a proximal adaptor and prototype and distal adaptors, while no significance between a distal and a prototype adaptor.

Simultaneous fit of the likelihood model to phoneme classification responses with and without adaptation

In the first experiment, we probed 'static-state' phoneme representations across a wide spectrum of sound stimuli. These data will allow us to evaluate the viability of the proposed model in terms of how accurately it captures the static-state relationship between physical property of sound stimuli and phoneme perception. The second experiment was designed to evaluate how capable the same model is of incorporating 'dynamic-state' phoneme representations, which refers to temporary changes in phoneme representation caused by temporal contextual modulation. Here we challenge the viability of the model by testing whether it can simultaneously (meaning with the same set of model parameters) capture the two sets of data, one associated with static-state phoneme representations and the other with dynamic-state phoneme representations. In addition, another challenge that the model has to address is to predict the two metrics of phoneme classification responses, CFs and RTs, again simultaneously.

For the data from the both experiments, the input to the model was the angular similarity value, θ, defined in radian space (Fig. 2b). At the sensory encoding stage, the model had 360 SN model neurons whose tuning functions were determined by the cyclic cosine function:

$$f_i(\theta) = ae^{b_{SN}(\theta - x_i)} + m \quad (5),$$

where *a,* b_{SN} , x_i and m are the height-scaling, width, peak location and baseline parameters, respectively, of tuning curve. All the other parameters except for b_{SN} were fixed such that the minimum and maximum values of fi are 0 and 1. In other words, the model assumed that the tuning curves of SNs are identical in shape and differ only in its preferred stimulus θ, such that p_i of the *i*th neuron equal to *i* deg in angular scale. Only one single model parameter, b_{SN}, was set to be free to adjust the width of the tuning curves of the entire population of SNs.

At the phoneme likelihood decoding stage, the model had three phoneme neurons with three different phoneme preferences /da/ ($SN_{/da/}$), /ba/ ($SN_{/ba/}$) and /ga/ ($SN_{/ga/}$). The weighting function for summation of SN j, w_j, was also model by the cyclic cosine function:

$$w_j\,(x) = ae^{b_j(x-p_j)} + m \quad (6),$$

where *x* is the preferred stimulus θ of a given SN and p_j is the center of weighting curve pea. All the other parameters except for b_j and p_j were fixed such that the sum of values of wj across the entire population of SNs is equal to zero. This 'zero-sum' constraint was adopted to implement a moderate level of tuned suppression. Since each of the three SNs has two free parameters, a total of 6 parameters were set to be free at the phoneme likelihood decoding stage.

At the evidence accumulation stage, the model translated the outputs of the previous stage into final model predictions, CF_M and RT_M with the following set of four parameters: t_r, a residual processing time reflecting afferent sensory processing plus efferent motion execution times; r_A, a mean rate of evidence accumulation; v_E, a variability in phoneme evidence; v_r, a variability in evidence accumulation rate. The last two parameters can be seen as terms representing levels of across-trial and within-trial

noise at the evidence accumulation stage. Note that we implemented these noise parameters at final stage for computational convenience. In principle, the model assumes that those two types of noise can arise anywhere along the hierarchy (probably across all of the three stages), e.g., across- and within-trial changes in firing rates in SNs, PNs and EANs.

The eleven parameters described so far were all common to the first (static) and second (dynamic) data sets, except for only the two parameters at the evidence accumulation stage, t_r, and r_A. The model assumed so because the residual time and accumulation rate can differ between the two experiments because the former can fluctuate across different daily sessions and the latter can be adjusted adaptively by listeners in reaction to different speed and accuracy requirements between the two experiments (e.g., weakened and ambiguous evidence after adaptation). The final parameter, ar – sensory adaptation rate–, which is unique to the second experiment, was introduced to implement adaptation-induced changes in response gain of SNs at the stimulus encoding stage. The model assumed that the degrees of adaptation for SNs are proportional to their preference strength to a given adapting stimulus. Thus the population profile of gain after adaptation is a scaled copy of the reciprocal of population responses to an adapting stimulus. The adaptation rate parameter, ar, determined the height of this post-adaptation gain profile.

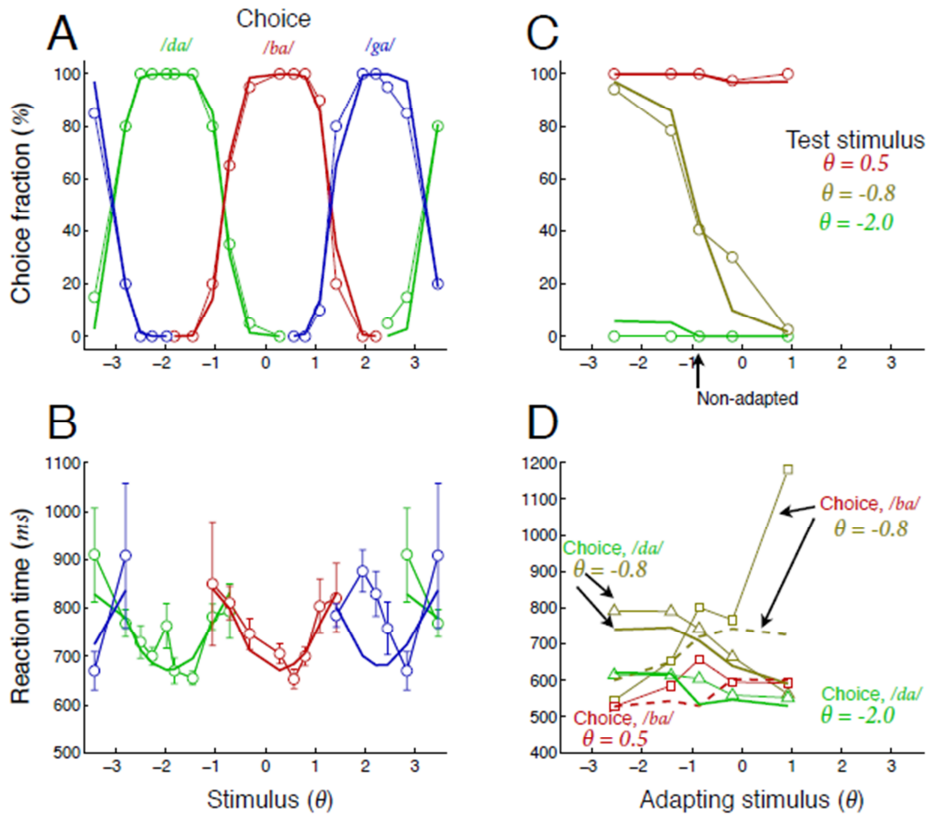


Figure10. Results of model fitting from one representative listener. Thick lines are the model predictions, and thin lines with symbols are the observed data. (a),(b) Results from Exp 1. Choice fractions (a) and reaction times (b) are plotted against the stimulus angles expressed in the radian scale. Different colors represent choices made by listeners. The error bars are standard errors of means. (c),(d) Results from Exp 2. Choice fractions (c) and reaction times (d) are plotted against the angles of adapting stimuli. Different colors represent three different test stimuli. There was no adapting stumulus for the data points denoted by the small vertical arrow. In (d), the thin lines with triangles and squares are observed RTs for choice /da/ and /ba/, respectively, whereas the thick solid and broken lines are model predictions of choice /da/ and /ba/, respectively.

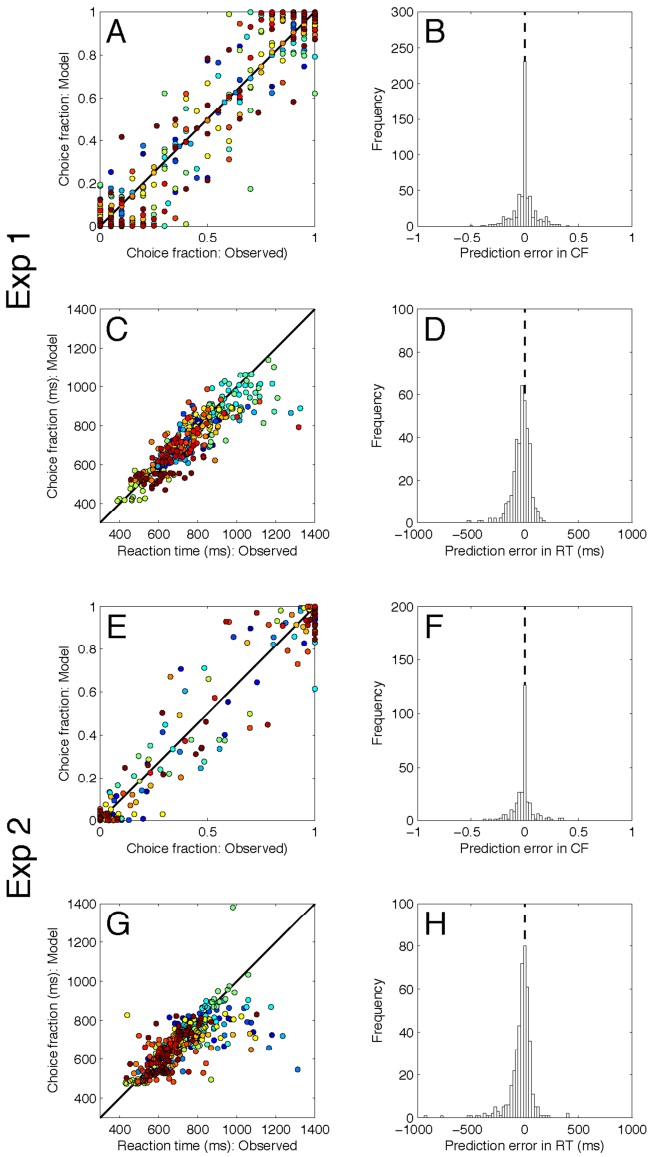


Figure11. Correlations between observed data and model predictions for the data merged across all sixteen listeners. (a-d) Resutls from Exp 1. (a),(c),(e),(g) The model predictions of choice fraction (a, e) and reaction times (c, g) are plotted against the observed data for corresponding stimuli. Colors represent individual listeners. (b),(d),(f),(h) The histograms show the number of data points for signed errors of model predictions.

With these 14 parameters set to be free for individual listeners, we fitted the model simultaneously to the two sets of data from the two experiments, each of which consisted of the two different metrics of phoneme classification responses, CPs and RTs. The model was fitted by the maximum likelihood method, in which likelihood distributions of CP and RT data were estimated from binomial process and standard error estimates, respectively. In average, 10 rounds of 500~1,000 iterations were enough to reach stable values of maximum likelihood.

As seen in Figure 10, in which the observed data (symbols with thin lines) and the model predictions (thick lines) from a representative listener were plotted together, the model successfully captured the CP and RT data from the two experiments simultaneously. Despite substantial amounts of individual differences, the model predictions accounted for a substantial fraction of variance in the observed data for the both types of metrics: the across-listener means (and standard deviations) of Pearson linear correlations were 0.974 (0.020) for CF in Exp 1, 0.577 (0.185) for RT in Exp 1, 0.980 (0.013) for CF in Exp 2, and 0.431 (0.265) for RT in Exp 2. As an alternative measure of goodness of model fit, we merged all the individual data points and the corresponding model predictions and computed the correlation between the two for each of the 2 experiments x 2 metrics conditions (Fig 11). The correlations were very high and significant ($p<10^{-10}$) in all the four conditions, 0.974 for CF in Exp 1; 0.872 for RT in Exp 1; 0.980 for CF in Exp 2; 0.648 for RT in Exp 2. The regression analysis also shows that the observed data and the model predictions were well matched in absolute value (mean data) without any noticeable presences of stimulus-regime-dependent prediction errors (non significant correlation between stimulus and model prediction errors).

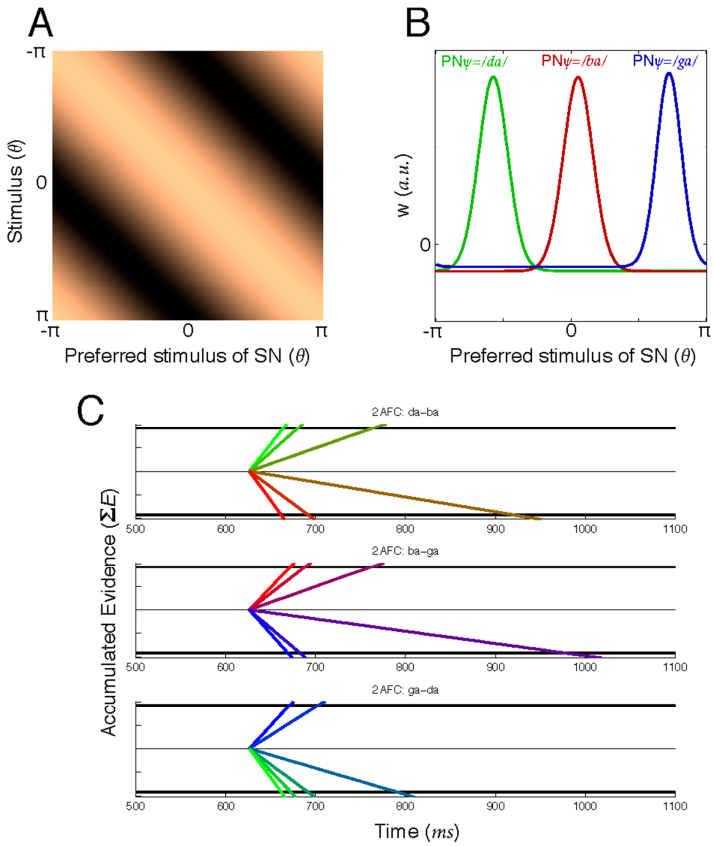


Figure 12. Illustration of key parameters of the model fit to the data for a representative listener. (a) Matrix of tuning curves of SNs at the sensory encoding stage. The brightness represents response amplitude to a given stimulus (y axis) for a given SN with a certain stimulus preference (x axis). (b) Weighting functions of PNs at the likelihood decoding stage. The green, red and blue curves are summation weights for PNs with preferences for phoneme /da/, /ba/ and /ga/, respectively. (c) Buildup of evidence in EANs at the evidence accumulation stage. The lines with different colors represent the noise-free temporal accumulation of decision evidence for different stimuli used in Exp1. Thus, the slopes of the lines represent the rates of evidence accumulation rate of the fitted model. The time point at which all of the lines begin to change simultaneous is the residual time (time taken for afferent sensory processing and for efferent motion execution). At each panel, the thick horizontal lines represent the bounds for decision. The three rows are for the three different blocks of trials, in which a specific pair of phoneme alternatives was discriminated.

Figure 12 visualizes internal components at the encoding and decoding stages of the fitted model for the same listener whose data were shown in Figure 10. By inspecting the relationship between the observed data and the fitted model parameters, we could gain a few insights how the model managed to capture phoneme classification responses common to all the individuals and those varying across individual listeners. First, the width of the sensory tuning curves at the encoding stage was quite broad in the majority of listeners (median b_{SN} = 0.364) whereas the width of summation weighting curves tended to be narrow and varied greatly across listeners (median b_{PN} = 7.021 for $PN_{/da/}$, 5.940 for $PN_{/ba/}$, 1.735 for $PN_{/ga/}$,). Second, the individual differences in category boundary location, which were quite substantial in the data from the first experiment, could be mostly explained by the variability in peak location of summation weighting curve. Third, previous studies reported that adaptation to a particular phoneme stimulus leads to a substantial repulsive shift of phoneme category boundaries (Samuel, 1986), as we clearly observed in the second experiment. The proposed model offers one plausible mechanistic-level account for this adaptation-induced repulsive shift of phoneme categories. According to the model, one simple scenario that adaptation occurs mainly at the sensory encoding stage and is inherited to the phoneme encoding stage, wherein the imbalance between two neighboring phonemes is produced. This imbalance would result in a distorted value of evidence, which leads to changes in CF and RT as well.

Discussion

Given a dearth of models for pre-lexical unit representation despite its importance in speech perception, we aimed at proposing a biologically plausible model that provides theoretical accounts for how pre-lexical units can be represented in the brain while capturing the key behavioral dynamics in actual human perception of pre-lexical units. We developed a system-level neural model for phoneme classification, with an assumption that phonemes are represented as probabilistic quantities in the brain. Our model is distinguished from and goes beyond previous models (McClelland & Elman, 1986; Norris et al., 2000)in several aspects.

First, the proposed model is plausible in a neural perspective. Our model not only specified a hierarchical set of probabilistic computations required for phoneme classification but also proposed the ways each of those computations can be achieved by the activity of a specific population of neurons in parallel. The model consists of three stages. The computation at the encoding stage is to calculate a set of individual likelihoods for acoustic features in incoming sound input that are relevant for distinguishing phonemes. The feature dimension crucial for distinguishing between stop-consonant syllables, which were target phoneme stimuli in the current study, is spectral-temporal dynamics in frequency harmonics during the initial 50 ms period after stimulus onset. Single-cell studies on animals reported that a substantial fraction of neurons in the primary auditory cortex has a spectral-temporal receptive field ('STRF') structure that is broadly tuned around a certain form of frequency modulation (Bandyopadhyay & Young, 2004; Mesgarani, David, Fritz, & Shamma, 2008). Our model posits that those neurons are likely candidates for operating the proposed computation at the encoding stage. As shown previously (Seung, H.S. & Sompolinsky, H 1993; Jazayeri & Movshon 2006) and here (Eq 1), the likelihoods can be computed from individual neurons' responses once their tuning functions are known. At the second stage occurs the most crucial computation, decoding likelihoods for given phonemes from individual sensory likelihoods that are available at the outputs of the computation

at the first stage. According to the model, this computation can be operated by weighted summation of presynaptic inputs at post-synaptic neurons. This 'weighted summation' at neural synapses is one of the most fundamental and ubiquitous circuits in the brain, evidenced by many sensory neural systems (Carandini, 2012; Rust, Mante, Simoncelli, & Movshon, 2006). Compared to other typical sensory computations, phoneme classification is unique in that its goal is not to estimate a certain physical quantity (e.g., orientation of an image contour for low-level vision), but to estimate the identity of a 'phoneme' that was intended by a speaker. This situation requires a listener's brain to perform a reverse engineering of the sound generation process performed by the speaker. The reverse engineering is equivalent to the 'weighted summation' computation at the second stage. Recent imaging studies reported the presence of neurons with selective preferences for phoneme stimuli in several areas located in the superior temporal cortex, which are believed to receive outputs of neurons in the primary auditory area, presumably including neurons with 'STRF' as proposed by our model (Obleser & Eisner, 2009b). The computation operated at the final stage is to form and integrate the evidence for making a discrete decision. This operation can be performed on the computational output at the phoneme likelihood decoding stage by selecting neurons (or neural populations) with preferences for phonemes relevant to an impending task. Previous single-cell studies (Gold & Shadlen, 2007) have been reporting that neurons in many cortical areas including superior colliculus, lateral infra-parietal area (LIP), frontal eye-field area exhibited evidence accumulation behavior. In addition, Yang and Shadlen (2007) convincingly demonstrated that responses of monkey LIP neurons integrate sequential probabilistic evidences (likelihood) for a binary decision. In summary, neural implementations of the computations at the core of our model are either supported by empirical observations or at least highly plausible given their ubiquitous presence in other sensory or cognitive brain regions.

Second, our model is well constrained by phoneme classification performance by human listeners. Using a small number of parameters, the model captured the dynamical changes in the two key aspects of phoneme classification behavior under various

listening conditions. In terms of behavioral aspects that were measure, our model offered joint predictions for the two key metrics of phoneme classification behavior: choice fraction and reaction time. In terms of contextual modulation, our model was capable of describing not only the static, categorical structure of phoneme classification as a function of physical stimulus property, but also the dynamic changes in classification on an identical stimulus when listeners were exposed to biased samples of sound stimuli for a prolonged period of time. Here, it should be underlined that this high degree of explanatory power was achieved with a single, constant set of parameters for the key computational elements of the model, including tuning curves of SNs, weighting curves of PNs and evidence weight and accumulation rate of EANs. Furthermore, the model was flexible enough to cover the wide range of individual differences in phoneme classification behavior.

Finally, from a theoretical perspective, our model is comprehensive in that it incorporates the cascade chains of essential information flows that start from *true* phonemes intended by speakers and end at phoneme estimates that can be used as inputs to the lexical system. Due to this comprehensive nature, the inspection of the way the model operates provides valuable insights regarding a few general features of human speech perception, and thus can guide future empirical neurophysiological or psychophysical studies by generating testable hypotheses and predictions. For example, the individual differences that were witnessed in our behavioral experiments are likely associated with listeners' idiosyncratic speech environments (e.g., regions with strong dialects). Our model points to the 'phoneme weighting functions' between SNs and PNs as neural correlates of those individual differences. This 'reverse engineering' hypothesis generates an intriguing possibility that phoneme neurons' response properties are readily modifiable, or alternatively new phoneme neurons evolve, by long-term exposures to a new sample of phoneme sounds generated by speakers.

References

Bandyopadhyay, S., & Young, E. D. (2004). Discrimination of voiced stop consonants based on auditory nerve discharges. *The Journal of neuroscience, 24*(2), 531-541.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review, 113*(4), 700-765. doi: 10.1037/0033-295X.113.4.700

Carandini, M. (2012). From circuits to behavior: a bridge too far? [Research Support, Non-U.S. Gov't]. *Nat Neurosci, 15*(4), 507-509. doi: 10.1038/nn.3043

Dayan, P., & Abbott, L. F. (2001). *Theoretical Neuroscience*. Cambridge: MIT Press.

Gold, J. I., & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience, 30*(1), 535-574. doi: 10.1146/annurev.neuro.29.051605.113038

Jazayeri, M., & Movshon, J. A. (2006). Optimal representation of sensory information by neural populations. *Nature Neuroscience, 9*(5), 690-696. doi: 10.1038/nn1691

Kohn, A., & Movshon, J. A. (2003). Neuronal adaptation to visual motion in area MT of the macaque. [Research Support, U.S. Gov't, P.H.S.]. *Neuron, 39*(4), 681-691.

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: dialects, idiolects, and speech processing. [Research Support, N.I.H., Extramural Research Support, U.S. Gov't, Non-P.H.S.]. *Cognition, 107*(1), 54-81. doi: 10.1016/j.cognition.2007.07.013

Lee, H. A., & Lee, S. H. (2012). Hierarchy of direction-tuned motion adaptation in human visual cortex. [Randomized Controlled Trial Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. *J Neurophysiol, 107*(8), 2163-2184. doi: 10.1152/jn.00923.2010

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1-86.

Mesgarani, N., David, S. V., Fritz, J. B., & Shamma, S. A. (2008). Phoneme representation and classification in primary auditory cortex. *The Journal of the Acoustical Society of America, 123*, 899.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*(03), 299-325.

Obleser, J., & Eisner, F. (2009a). Pre-lexical abstraction of speech in the auditory cortex.

[Research Support, Non-U.S. Gov't]. *Trends Cogn Sci, 13*(1), 14-19. doi: 10.1016/j.tics.2008.09.005

Obleser, J., & Eisner, F. (2009b). Pre-lexical abstraction of speech in the auditory cortex. *Trends in cognitive sciences, 13*(1), 14-19.

Perkell, J. S., & Klatt, D. H. (1986). *Invariance and variability of speech processes.* Hillsdae, NJ: Lawrence Erlbaum.

Pouget, A., Dayan, P., & Zemel, R. S. (2003). Inference and computation with population codes. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Research Support, U.S. Gov't, P.H.S. Review]. *Annu Rev Neurosci, 26*, 381-410. doi: 10.1146/annurev.neuro.26.041002.131112

Rust, N. C., Mante, V., Simoncelli, E. P., & Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nature Neuroscience, 9*(11), 1421-1431.

Samuel, A. G. (1986). Red herring detectors and speech perception: In defense of selective adaptation. *Cognitive Psychology, 18*(4), 452-499.

Schwartz, O., Hsu, A., & Dayan, P. (2007). Space and time in visual context. [Research Support, Non-U.S. Gov't Review]. *Nat Rev Neurosci, 8*(7), 522-535. doi: 10.1038/nrn2155

요약(국문초록)

말소리 인지에서 어휘 전 단계과정의 중요성에도 불구하고 현재까지의 모델들은 어휘단계나 그 이후 단계의 언어인지과정에 초점을 맞추어 왔다. 어휘 전 단계에 대한 기존 체계의 부족을 깨닫고 우리는 이 연구에서 시스템 수준에서 음소 분류의 신경학적 모델을 제시하고자 한다. 우리 모델의 핵심은 음소가 뇌에서 확률 값인 '우도'로써 표현된다는 것이다. 또한 모델은 우리 뇌에서 잘 알려진 기본적인 계산활동인 '감각 부호화', '우도 출력', '증거축적' 세가지를 소리 자극의 입력으로부터 말소리 인지에 이르는 순차적 위계 신경 프로세스를 통해 잘 설명하고 있다. 모델의 첫 번째 단계에서는 음소 구별에 필요한 물리적 특징의 조율 곡선을 가진 감각 뉴런들이 각 특징들의 존재 여부에 대한 우도를 계산한다. 다음 단계인 음소 뉴런에서는 각 음소 별 가중치 곡선을 이용해 감각 뉴런들의 입력을 합하여 특정 음소의 우도를 계산한다. 마지막 단계인 증거 축적 뉴런에서는 과제에 적합한 음소 뉴런들의 입력을 통합하여 시간에 따른 증거들을 계산, 축적하여 특정 음소를 선택하게 한다. 증거 축적에 따른 선택 메커니즘은 음소뉴런에서 오는 확률적인 정보로부터 개별적 음소선택 출력이 가능하게끔 한다. 이러한 해석과정은 다양한 청취상황 내의 음소판단과제에서 얻은 선택과 반응시간의 데이터를 이용해 우리의 모델의 적합성을 알아보는 것을 가능하게 하였다. 적은 수의 변수를 사용한 우리의 모델은 일반적인 음소판단 과제 상황 뿐만 아니라 선택적 순응이 유도된 동적 상황에서의 같은 과제에 대한 반응 패턴까지도 예측하였다. 더 나아가 우리의 모델은 음소 구별 행동의 개인차까지도 반영할 정도로 융통성이 있음을 보였다. 우리의 모델은 신경 생리학적 실험이나 행동적 실험으로 확인 가능한

가설과 예측을 제공함으로써 말소리 인지과정을 초기 단계를 이해하는 기틀을 마련하였다.

주요어: 음소 분류화, 지각 판단, 음소 인지, 감각 부호화, 우도 출력, 신경학적 모델
학번: 2010-24020