


ORIGINAL ARTICLE

# Performance pay and non-native language comprehension: Can we learn to understand better when we're paid to listen?

Chasen Afghani<sup>1</sup>, Melissa M. Baese-Berk<sup>1,2</sup>  and Glen R. Waddell<sup>3,4</sup>

<sup>1</sup>Department of Linguistics, University of Oregon, Eugene, OR, USA, <sup>2</sup>Department of Linguistics, University of Chicago, Chicago, IL, USA, <sup>3</sup>Department of Economics, University of Oregon, Eugene, OR, USA and <sup>4</sup>IZA Bonn, Bonn, Germany

**Corresponding author:** Melissa M. Baese-Berk; Email: [mbaesebe@uoregon.edu](mailto:mbaesebe@uoregon.edu)

(Received 11 July 2022; revised 2 June 2023; accepted 12 June 2023)

## Abstract

Non-native speech is difficult for native listeners to understand. While listeners can learn to understand non-native speech after exposure, it is unclear how to optimize this learning. Experimental subjects transcribed non-native speech and were paid either a flat rate or based on their performance. Participants who were paid based on performance demonstrated improved performance overall and faster learning than participants who were paid a flat rate. These results suggest that exposure alone is not sufficient to optimize learning of non-native speech and that current models of this process must be revised to account for the effects of motivation and incentive.

**Keywords:** Adaptation; incentive; non-native speech; pay for performance; speech perception

## Introduction

Communication between native and non-native speakers of a language is increasingly common in our globalized society. Within the United States, nearly 10% of school-aged students are classified as English Language Learners and, in some states, almost one-in-five students are English Language Learners (e.g., 19.2% in California and 18% in Texas; Department of Education, 2020). Outside of the educational system, 20.5% of the U.S. population speaks a language other than English at home, and nearly 40% of that population reports that they speak English at a level below “very well” (American Communities Survey, 2018). English serves as a “lingua franca” or common language of communication in many contexts internationally, including business and trade (Brutt-Griffler, 2005; Rogerson-Revell, 2007). There are nearly double the number of non-native speakers of English globally than native English speakers, a gap that continues to grow (Kachru, 1986). Further, conversation in English often occurs among parties who do not share a

© The Author(s), 2023. Published by Cambridge University Press.. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

native language background. Previous research has demonstrated that non-native speech is more difficult for native speakers to understand than native speech (Munro and Derwing, 1995). However, individuals are able to improve their comprehension of accented speakers with relatively limited exposure (Baese-Berk *et al.*, 2013; Bradlow and Bent, 2008). Investigations of this adaptation have focused exclusively on properties of speech that underly this adaptation (e.g., how the acoustic properties of speech modulate adaptation; Xie & Myers, 2017) or on cognitive factors that may impact adaptation (e.g., working memory; Rönnerberg *et al.*, 2013).

Here, we examine how direct incentives to perform modulate adaptation, or learning, and thereby contribute along multiple dimensions. For example, to an established literature we offer support for the external validity of existing results. Namely, existing laboratory studies have found that subjects adapt to non-native speech (Baese-Berk *et al.*, 2013; Bradlow and Bent, 2008), but have based their support for this on experimental environments that are without direct performance incentives. Yet, outside of the laboratory, very many of the relevant environments—those in which conversations occur among parties who do not share a native language background—are fundamentally incentivized. For example, in professional or academic environments, there is clear benefit for listeners to both accurately and efficiently communicate. That is, in natural communication situations where participants may not share language backgrounds, one could imagine that participants in these conversations would be inherently incentivized to communicate well—both in terms of their own speech clarity but also in exerting more listening effort when they may be less familiar with a specific talker or their accent. However, this incentive is indirect, rather than a direct performance incentive.

This experiment measures whether directly incentivizing participant's performance by attaching monetary rewards to how well they identify non-native speech impacts their performance and whether adaptation to unfamiliar speech is accelerated. In Section 2, we provide some context and background to the larger literatures we implicate. In Section 3, we describe the experiment and summarize the data-generating process, before discussing our empirical analysis in Section 4. We follow the analysis with a brief discussion of the policy implications in Section 5.

## Background

### *Speech perception*

Speech perception is a notoriously difficult task, requiring listeners to generalize over substantial variation within and across speakers in order to successfully perceive the intended message from a speaker. That is, multiple acoustic signals could map onto a single word—the way Person A produces the word “cat” will result in a different speech signal than the way Person B produces the same word. Listeners are, generally, extraordinarily good at handling this variability and typically understand speech with relatively little effort. However, in some circumstances, speech perception is less successful. For example, perceiving speech in noisy situations is more challenging than understanding speech in quiet (Cherry, 1953). Similarly, understanding speech from an unfamiliar talker, especially from a

speaker with an unfamiliar accent, is more challenging than listening to familiar speakers or accents (e.g., Nygaard, Sommers, & Pisoni, 1994; van Wijngaarden, 2001).

The hallmark case of unfamiliar speech is non-native speech. Non-native speech deviates from native speech on a variety of dimensions, but one of the most salient dimensions is the accent (or features of pronunciation) that differs from native speech. These pronunciation differences may occur at the segmental level (i.e., individual speech sounds) or at a suprasegmental level (overall pitch, rhythm, etc.). Taken together, these features create a distinct acoustic profile that is often challenging for native listeners to understand. Substantial previous work has investigated the sources of these challenges and has asked why some listeners succeed more than others on the task of understanding non-native speech. Cognitive factors including vocabulary size have been shown to predict a listener's ability to understand non-native speech (Banks et al., 2015; Bent et al., 2016; McLaughlin et al., 2018). Further, in a matched-guise task where the same speech sample from a native English speaker is matched with either an Asian face or a Caucasian face, listeners report that speech is more accented when paired with the Asian face (Rubin, 1992). Further, listeners transcribe speech more accurately when the race of the speaker matches the accent of the speech (e.g., a Chinese face and Chinese-accented English, McGowan, 2015). In addition to these factors, attitudinal factors also impact perception (see, e.g., Kutlu et al., 2022a, 2022b). Listeners with more negative attitudes toward non-native speakers report the speech as being more challenging to understand, even if they are equally able to transcribe the speech (Sheppard, Elliott, & Baese-Berk, 2017).

While it is clear that a variety of factors impact baseline perception of non-native speech, it is also the case that listeners are able to improve their perception of non-native speech with some practice. Some previous studies suggest that initial adaptation can be relatively quick (i.e., within a few sentences, Clarke and Garrett, 2004), other work has demonstrated that longer periods of exposure (e.g., 30 minutes of training over the course of two days) result in significant improvements to perception of non-native speech (Bradlow and Bent, 2008). Listeners can improve their perception of a specific accented talker, on a variety of talkers from a single accent background, or on talkers from a variety of accent backgrounds, depending on the speakers they are exposed to during training (e.g., Baese-Berk, Bradlow, & Wright, 2013; Bradlow & Bent, 2008; Sidaras, Alexander, & Nygaard, 2009). Thus, while we know that listeners can improve at understanding unfamiliar, accented speech, it is unclear what factors impact this adaptation.

The consequences of having a non-native accent extend far beyond communication specifically, as non-native speakers face myriad biases (Gluszek and Dovidio, 2010). Individuals with non-native accents are often viewed as less employable than native speakers (Carlson and McHenry, 2006) and are less likely to be recommended for a promotion or to receive entrepreneurial investments (Huang, Fridleger, & Pearce, 2013). Further, non-native speakers may be judged as less credible (Lev-Ari and Keysar, 2010), a bias that might emerge in early childhood (Kinzler, Corriveau, & Harris, 2011). These judgments are often tied to judgments about the speaker's language and challenges for the listener when understanding that speech. Indeed, some work suggests that listeners who have more experience with speakers from a variety of language backgrounds impact both a listener's

attitude about the speaker and their ability to understand the speaker's speech (Kutlu *et al.*, 2022b). Therefore, it is critically important to understand how listeners can best improve their ability to perceive non-native speech.

### ***Pay for performance***

That incentives matter to human behavior is so foundational to economics that many introductory lectures often start with lessons from history. Many modern textbooks (Cowen and Tabarrok, 2018) retell stories of convict ships in the 1700s, for example, when the British government paid sea captains to take felons to Australia. Yet, many would not survive the voyage. In response, the government tried to fix the problem with myriad solutions (e.g., mandating that captains bring medical personnel on the voyage or requiring them to bring lemons to prevent scurvy). However, nothing worked until the pattern of paying for each prisoner that walked on the ship in Great Britain was abandoned and replaced by a system that paid captains for each prisoner that walked off the ship in Australia. The change in incentives aligned the self-interest of the captains with the self-interest of the convicts, and the captains responded to the incentives.

In the most general terms, an incentive is anything that motivates a person to do something. As we approach our research question, it will also serve well to distinguish two types of incentives. Namely, intrinsic incentives come from within—a person with an intrinsic motivation wants to do something for its own sake, without an outside pressure or reward. The contributions to what intrinsically incentivizes individuals are many and varied, from feeling personal fulfillment and satisfaction from doing certain things, or from learning a new skill just for the fun of it. On the other hand, extrinsic incentives involve providing a material reward for accomplishing a task (a positive incentive) or threatening a punishment for failure to do so (a negative incentive).

Thus, in the absence of extrinsic incentives—we will use money to incentivize subject performance in our experiment—it is not the case that experimental subjects are then without incentive at all—this is true here and in any laboratory of human subjects. Rather, our design will allow us to “difference out” the intrinsic incentives that are likely to be common to both treated and control subjects and leave the extrinsic incentive provided only to the treated group as the implicated mechanism explaining the difference in performance between the two groups.<sup>1</sup>

That human subjects respond to incentives is well established, across a variety of environments (e.g., Haley, 2003; Lazear, 2000; Seiler, 1984; Shearer, 2004). The role of motivation, a concept closely related to incentive, has been implicated in previous theories of speech perception in challenging listening situations, namely in the Framework for Understanding Effortful Listening (FUEL; Pichora-Fuller *et al.*, 2016). This model integrates multiple factors ranging from cognitive to interpersonal that may impact how a person understands speech in challenging listening situations, such as listening to a speaker with an unfamiliar accent. The present study builds on this proposal by explicitly manipulating extrinsic incentives to participants to investigate this component of a broader construct of motivation. Further, while some previous work has investigated the role of monetary reward on listening effort during speech perception in noise (e.g., Koelewijn *et al.*, 2018, 2021),

this study directly investigates how monetary reward impacts both performance and, crucially, improvement in performance over time.

## **Method**

### ***Positionality statement***

The three authors of this work are affiliated with the University of Oregon. Chasen Jaleh Afghani identifies as an Iranian-American woman. She identifies as speaking standardized American English, having Farsi as a heritage language and Spanish as an L2. Her interest in linguistics production and perception stems from her background. Melissa Michaud Baese-Berk identifies as a white, educated, cisgender woman, and identifies as speaking a standardized form of American English. While she has also identified as a learner of a variety of languages and has spent time as a “non-native” speaker in environments where those languages are dominant (e.g., a Spanish learner living in Spain), she has spent most of her life in environments where her language variety is the dominant form across institutions of power. She uses behavioral methods to investigate speech perception, speech production, and language learning across a wide array of language varieties. Glen Waddell identifies as a behavioral social scientist, data scientist, and economist. He speaks a standardized form of American English and has spent most of his life in environments where his language variety is the dominant form.

### ***Participants***

In January 2020, we recruited adult native English speakers from the student population at University of Oregon. All individuals were adults (ages 18–28 (mean = 22; sd = 2.2); 30 female, 20 male). We recruited participants using a recruitment flyer sent by e-mail and shown in various classes. Subjects were chosen to participate in the study if they self-identified as native, monolingual English speakers with limited experience with non-native accented speech (i.e., did not have a family member, close friend, or roommate who is a non-native speaker of English). Further, no subject reported having a history of speech, language, or hearing disorder; however, demographic characteristics of participants were not collected. A total of 50 subjects were recruited—25 were chosen at random to experience the treatment regime and 25 were the control.

### ***Methods***

Participants completed a sentence transcription task (i.e., an intelligibility task) and a questionnaire. All tasks were administered via PsychoPy (Peirce, 2007). Listeners heard sentences over headphones. At the conclusion of each sample, they were asked to type exactly what they heard. Listeners also completed a questionnaire about their experience with other languages, and accents of English.

Participants were assigned to one of two groups. In the treatment group, subjects were rewarded for their ability to correctly identify the words they have been presented with (i.e., intelligibility). Subjects in this group were paid a \$9 “show-up”

fee plus \$2 for each word they correctly identified in one of the (104) samples. Subjects received payment for one such sample, which was determined randomly, with equal chance of it being any of the samples they experienced. In the end, the mean payment to subject in the treatment group was \$15.88, with minimum and maximum payments of \$9 and \$19. In the control group, subjects were paid a flat fee of \$14 for their participation and no other direct incentive was given.<sup>2</sup> Experiments were performed in the Spoken Language Research Laboratories, where it took participants roughly 60 minutes to complete all experimental tasks.

### **Materials**

Stimuli for the experiment were drawn from the Hearing in Noise Test subsection (Nilsson, Soli, & Sullivan, 1994) of the Archive of L1 and L2 Scripted and Spontaneous Transcripts and Recordings (ALLSSTAR corpus; Bradlow, Kim, & Blasingame, 2017), a publicly available corpus of native and non-native speech.<sup>3</sup> Stimuli were chosen from six native Mandarin talker, three men and three women. In Table 1, we reproduce the 104 target sentences employed in this experiment. Sentences are between five and seven words long. All subjects experienced the same 104 sentences, though their order was randomized across subjects. Following previous work, these stimuli were embedded in speech-shaped noise at a 1:1 (i.e., 0 dB) signal-to-noise ratio to avoid ceiling effects (Bradlow and Bent, 2008). Each response was scored for the number of words correctly transcribed by the listener using Autoscore (Borrie, Barrett, & Yoho, 2019). Words had to be entirely correct and partial credit was not given.

### **Results**

We begin by describing the pattern of results we observe in the data. In Figure 1, we plot the average number of correctly identified words in each of the 104 target sentences. We separately identify (in orange) the average performance among subjects in the treatment group ( $n = 25$ ), and (in blue) the average performance among subjects in the control group ( $n = 25$ ). We also fit both groups to a third-order polynomial. Here, we first see a clear suggestion that there is a level increase in performance among the treated subjects—those who were given extrinsic monetary incentives to understand speech. However, the implied shape parameters also suggest a different pattern of learning emerges in the treatment and control groups—across the order of sentences, treated subjects not only start ahead, but *gain* over control subjects. Following these observations, we fit a series of regression models using R. We describe our statistical analyses of the data in detail below.

#### ***Do people listen better with incentive to do so?***

In Table 2, we report estimates from a series of models. In each, our objective is to measure the effect of monetary incentive on the number of words identified correctly in each of 104 target sentences presented to subjects—these will form our baseline specifications. Specifically, we model responses as:

**Table 1.** Target sentences. *Notes:* Stimuli for the experiment were sentences from the Hearing in Noise Test subsection (Nilsson et al., 1994) of the Archive of L1 and L2 Scripted and Spontaneous Transcripts and Recordings (Bradlow et al., 2017)

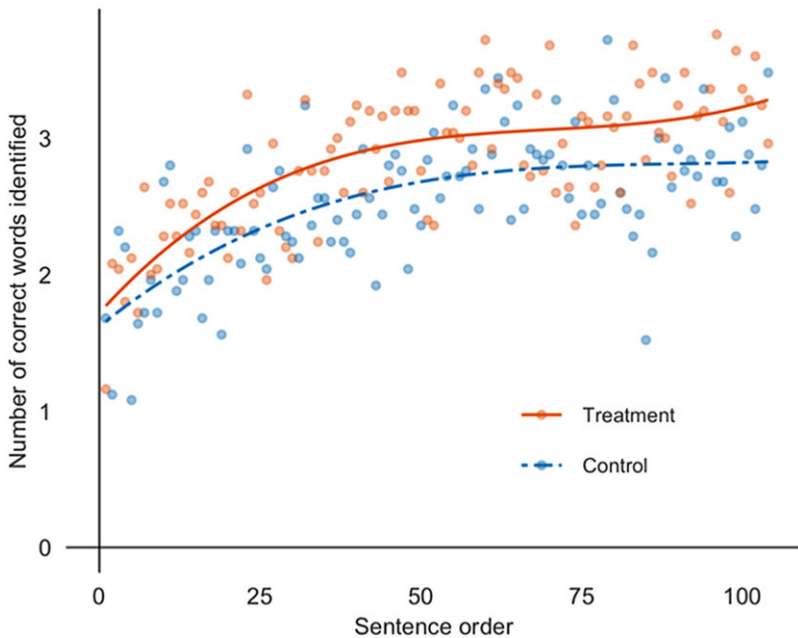
The y waited for an hour.	The dog sleeps in a basket.	The janitor swept the floor.
The dog is chasing the cat.	The food is expensive.	They heard a funny noise.
The shoes were very dirty.	The girl played with the baby.	She lost her credit card.
The tall man tied his shoes.	The bananas are too ripe.	The man is painting a sign.
They're shopping for school clothes.	The milk is in a pitcher.	They found his brother hiding.
The man is painting a sign.	The car is going too fast.	The road goes up a hill.
A letter fell on the floor.	The jelly jar was full.	The two children are laughing.
The bus leaves before the train.	The family bought a house.	The matches are on a shelf.
The firetruck is coming.	The football game is over.	The dog came home at last.
The children are walking home.	He grew lots of vegetables.	It's time to go to bed.
Rain is good for trees.	She cut the steak with her knife.	She stood near the window.
Sugar is very sweet.	The dog came home at last.	The towel fell on the floor.
The teapot is very hot.	They are coming for dinner.	The young people are dancing.
The kitchen clock was wrong.	The apple pie is good.	The dog played with a stick.
A towel is near the sink.	The milkman drives a small truck.	The mailman brought a letter.
The children helped their teacher.	There was a bad train wreck.	Someone is crossing the road.
Mother got a sauce pan.	Somebody stole the money.	The match boxes are empty.
A boy fell from the window.	Flowers can grow in the pot.	The ball bounced very high.
The shirts are in the closet.	He hung up his raincoat.	The boy did a handstand.
The baby wants his bottle.	School got out early today.	The paint dripped on the ground.
They're running past the house.	Her coat is on a chair.	The table has three legs.
The picture came from a book.	The lady washed the shirt.	They are drinking coffee.
The baby slept all night.	The boy ran away from school.	The new road is on the map.
The fire was very hot.	The boy is running away.	Mother read the instructions.
They sat on a wooden bench.	Father forgot the bread.	The cat lay on the bed.
The girl is washing her hair.	The bag fell off the shelf.	He's skating with his friend.
The child drank some fresh milk.	She found her purse in the trash.	The milk was by the front door.
They knocked on the window.	She took off her fur coat.	The mother heard the baby.

(Continued)



**Table 1.** (Continued)

The team is playing well.	The kitchen window was clean.	The cow is milked every day.
The baby has blue eyes.	The football hit the goalpost.	The book tells a story.
They took some food outside.	Swimmers can hold their breath.	Mother cut the birthday cake.
The waiter brought the cream.	They carried some shopping bags.	The policeman knows the way.
The ball broke the window.	They called an ambulance.	He climbed up the ladder.
The truck drove up the road.	The painter uses a brush.	The family likes fish.
They had two empty bottles.	They watched the scary movie.	



**Figure 1.** Mean Performance Across Target Sentences, by Treatment Status Here, We Plot the Mean Number of Correctly Identified Words in Each of the 104 Target sentences, Separately for Treatment Subjects ( $n = 25$ ) and Control Subjects ( $n = 25$ ).

$$\text{Number Correct}_{is} = \alpha + \mathbb{1}(\text{Treated}_i = 1) + f(\text{Sentence Order}_{is}) + \delta_s + \epsilon_{is}$$

where  $\text{Number Correct}_{is}$  captures the number of words subjects  $i$  correctly identify in target sentences  $s$ , and  $\mathbb{1}(\text{Treated}_i = 1)$  captures the treatment status of  $i$ . Subjects are known to learn with experience (Bradlow and Bent, 2008)—we model the systematic component of learning in  $f(\text{Sentence Order}_{is})$ . As subjects experience target sentences in random order, there is variation across subjects when a particular



**Table 2.** Estimates from six model specifications asking if performance incentives increase the number of words subjects correctly identify. In all specifications, we control for any systematic difference in the average performance on target sentences with target-sentence fixed effects. In (1)–(3), we allow subject performance to vary across order with a third-degree polynomial. In (4)–(6), we instead absorb any differences in the average performance by question order (i.e., the first, second, and third questions). Standard errors are reported in parentheses (i.e., \*\*\* 1%, \*\* 5%, and \* 10%)

	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.294**	0.294**	0.294**	0.294**	0.294**	0.294**
	(0.138)	(0.042)	(0.137)	(0.139)	(0.042)	(0.138)
Sentence FEs	✓	✓	✓	✓	✓	✓
Sentence order (cubic)	✓	✓	✓			
Sentence-order FEs				✓	✓	✓
Clustered SE	subject	sentence	subject and sentence	subject	sentence	subject and sentence
Control-group mean	2.54	2.54	2.54	2.54	2.54	2.54
Impact (%)	11.58	11.58	11.58	11.58	11.58	11.58
Effect size	0.15 sd	0.15 sd	0.15 sd	0.15 sd	0.15 sd	0.15 sd
Observations	5,200	5,200	5,200	5,200	5,200	5,200
Adjusted $R^2$	0.501	0.501	0.501	0.501	0.501	0.501

**Table 3.** Model specifications investigating whether performance improves more on longer sentences. In all specifications, we estimate standard errors allowing for clustering at the subject level, which we report in parentheses (i.e., \*\*\* 1%, \*\* 5%, and \* 10%)

	Full sample (1)	4 words (2)	5 words (3)	6 words (4)	7 words (5)
Treated	0.294** (0.139)	0.215** (0.126)	0.220 (0.135)	0.380** (0.160)	0.908** (0.275)
Sentence-order FEs	✓	✓	✓	✓	✓
Sentence FEs	✓	✓	✓	✓	✓
Control-group mean	2.54	1.96	2.16	3.25	3.08
Impact (%)	11.58	10.96	10.18	11.69	29.48
Effect size	0.15 sd	0.15 sd	0.15 sd	0.15 sd	0.15 sd
Observations	5,200	450	2,800	1,750	200
Adjusted $R^2$	0.501	0.52	0.43	0.516	0.198

target sentence is drawn. As such, throughout our analysis we will control for unobservable time-invariant heterogeneity specific to the 104 individual target sentences in  $\delta_s$ . As the treatment varies at the subject level, the anticipated level at which errors will cluster is with subjects. That said, in Table 2, we report estimated standard errors allowing for clustering at the subject level, at the sentence level, and at the subject + sentence level. Inference is not sensitive to this distinction.

In estimating (1), we are identifying the effect of monetary incentive as measured by the average difference (across the 104 target sentences) in the performance of the treated subjects on average and the control subjects on average.<sup>4</sup> However, we approach the modeling of the number of words subjects correctly identify in two distinct ways. In the first three columns of Table 2, we allow outcomes to change according to a third-order polynomial—that is, we estimate the effect of treatment having fit outcomes to  $f(\text{Order}_{is}) = \beta_1 \text{Order}_{is} + \beta_2 \text{Order}_{is}^2 + \beta_3 \text{Order}_{is}^3$ . In columns (4) through (6), we instead absorb any differences in the average performance on each of the 104 questions-orders (i.e., the first, second, third questions).<sup>5</sup> This non-parametric approach is less restrictive than to assume a cubic functional form to learning, yet, across both approaches we see similar point estimates, and only slightly different confidence intervals.

Treated subjects significantly outperform control subjects—we find 11.6% higher performance with monetary incentive (i.e., 0.294 additional words identified correctly in the average sentence). Incentivizing performance at a rate of \$2 per word increased average performance among treated subjects the equivalent of 15% of a standard deviation (i.e.,  $0.15\sigma$ ). In all cases, we reject that the average number of words correctly identified among treated subjects is equal to that among control subjects. In subsequent tables, we will adopt the specification of Column (4) of Table 2 as our preferred model—this represents the most conservative approach to inference, where we include sentence-order fixed effects, and estimate standard errors that allow for errors that may be correlated (across sentences) within subjects.

**Table 4.** Model specifications for average treatment/control differences across the order of target sentences to ask whether learning improves with incentives to perform. In all specifications, we estimate standard errors allowing for clustering at the subject level, which we report in parentheses (i.e., \*\*\* 1%, \*\* 5%, and \* 10%)

	First 15 (1)	First half (2)	Full Sample (3)	Last half (4)	Last 15 (5)
Treated	0.201	0.277**	0.294	0.312	0.330**
	(0.139)	(0.126)	(0.135)	(0.135)	(0.130)
Sentence-order FEs	✓	✓	✓	✓	✓
Sentence FEs	✓	✓	✓	✓	✓
Control-group mean	1.94	2.27	2.54	2.81	2.85
Impact (%)	10.34	12.2	11.58	11.09	11.58
Effect size	0.11 sd	0.14 sd	0.15 sd	0.16 sd	0.18 sd
Observations	800	2,600	5,200	2,600	800
Adjusted $R^2$	0.485	0.507	0.501	0.472	0.504

In Table 3, we demonstrate the robustness of the experimental results to the number of words in each sentence. As our preferred specification will include sentence-level fixed effects, we do not fear that unobserved heterogeneity across sentences drives treatment-effect estimates in Table 2. However, the opportunity for treated subjects to outperform control subjects may still differ with sentence length. Indeed, we see the largest gaps between treated and control subject on the seven-word sentences—subjects facing monetary incentives to correctly identify 29% more words ( $0.41\sigma$ ). As longer sentences may better reflect the realities of language and communication in the field, we see this increase in effect size as an indication that the potential improvements in performance we document in the laboratory are suggestive of meaningful improvements externally.

This methodological approach differs from that of previous studies (Baese-Berk et al., 2013) that have used a “training and test” approach to examining learning. In earlier experiments, two groups of participants are compared: one group who has been exposed to non-native speech during training and another which has been exposed to native speech during the training period. That is, both groups have experience with the task and with the laboratory setting before their performance is examined. As such, all participants are tested on novel talkers, and performance is only assessed at test. In these experiments, participants who have been previously exposed to non-native speech typically identify 0.31 additional words than those without the earlier exposure, or roughly 16% of a standard deviation in performance. These magnitudes are similar to those in the present study.

### ***Do incentives to perform induce faster learning, too?***

In Table 4, we stratify our baseline results by the order of sentences.<sup>6</sup> We first estimate the model of Equation (1) on a sample we restrict only to the first-15 sentences (Column 1). Here, treated subjects correctly identify 0.201 more words on

**Table 5.** Model specifications asking if performance incentives induce treatment and control groups onto different learning trajectories. In all specifications, we estimate standard errors allowing for clustering at the subject level, which we report in parentheses (i.e., \*\*\* 1%, \*\* 5%, and \* 10%)

	Treated (1)	Control (2)	Treated (3)	Control (4)
Order	0.010***	0.010***	0.074***	0.044***
	(0.002)	(0.002)	(0.011)	(0.009)
Order <sup>2</sup>			-0.001***	-0.001***
			(0.0003)	(0.0002)
Order <sup>3</sup>			0.00001***	0.00000
			(0.00000)	(0.00000)
Sentence FEs	✓	✓	✓	✓
Control-group mean	2.84	2.54	2.84	2.54
Observations	2,600	2,600	2,600	2,600
Adjusted $R^2$	0.529	0.458	0.538	0.463

average, relative to control subjects, which is equivalent to roughly 11% of a standard deviation in performance. Over the first 52 of the 104 sentences (Column 2), the gap between treated and control subject increases to 0.277 additional words ( $0.14\sigma$ ). In Column (3), we replicate our preferred specification on the full sample—there, the gap in the pooled model is 0.294 additional words ( $0.15\sigma$ ). As a general rule, treated subjects correctly identify differentially more words later in the experiment than they do early in the experiment, consistent with performance incentives not only increasing performance but inducing more-efficient learning.

This is further evidenced as we discard sentences experienced early in the experiment. For example, in Column (4), we restrict the sample to the last half of sentences, where the gap between treated and control subjects is higher still—to 0.312 additional words, or roughly 16% of a standard deviation in performance. In the last-15 sentences, the gap is also highest, increasing to 0.330 additional words, or roughly 18% of a standard deviation (Column 5). In the end, performance incentives increase performance by roughly 64% more in the last-15 sentences than in the first-15 sentences.

In Table 5, we offer a different approach to identifying the dynamics of learning in the treated and control groups. In columns (1) and (2), we separately fit outcomes to a “linear-learning” restriction.<sup>7</sup> Under such a restriction, we see no evidence of differential learning across treatment and control groups. Point estimates on their linear slopes are statistically indistinguishable. However, as was suggested in Figure 1, relaxing the linear restriction on learning in favor of a “cubic learning” technology reveals a richer story.<sup>8</sup> Not only do treated subjects correctly identify more words generally, they learn more quickly early and again late. Cubic learning cannot be rejected in either treated or control subjects. However, all three components are significantly different across treatment and control groups—in particular, the linear and cubic components are significantly more positive among those subjects facing direct incentives to identify words correctly. Performance

incentives increase subjects' ability to correctly identify words spoken and increase the rate of learning. It's as though learning is itself more productive with direct incentive.

## Discussion

Learning to listen to non-native speech has previously been understood as being an issue of exposure. The more non-native speech a listener has heard the better they are able to understand new speech from new talkers (Baese-Berk et al., 2013; Bradlow and Bent, 2008). The results of the current study offer a critique of those accounts along two dimensions. First, incentive alone results in a shift to initial performance. That is, individuals who are told they will be paid more based on their performance begin the experiment at a higher level of performance than individuals who are told they will be paid a flat rate, suggesting some aspects of performance may not be tied to exposure at all, but may instead be tied to motivation. Second, individuals who are incentivized to perform well on a task demonstrate more robust learning during the course of exposure than individuals who are not. Together, the results suggest that exposure alone does not result in the most robust learning. Instead, it appears that motivation, as indexed here by monetary incentive, can further improve learning, above and beyond exposure alone. Instead of theories of learning that rely primarily on exposure and modulating factors of this exposure, we must provide explanations for learning that are more nuanced, including exposure, motivation, and other social or attitudinal factors. Some prior studies have examined whether listeners process accented speech differently depending on their expectations about that speech signal (e.g., the source of the accent) (Lev-Ari, 2015; McGowan, 2015), and whether these expectations can impact how listeners are able to adapt to this speech (Vaughn, 2019) which also suggests that exposure alone cannot sufficiently account for improvement in perception of non-native speech. To our knowledge, no prior work has examined whether listener's attention and motivation to adapt to non-native speech can be explicitly modulated by attaching performance to a reward.

Performance incentives have been used as a proxy for motivation or encouraging shifts in attentional resource allocation in other auditory tasks. For example, increased motivation (also through providing monetary incentive) can improve performance on some auditory tasks for both hearing-impaired and normal-hearing listeners (Mirkovic et al., 2019). Interestingly, physiological measures (e.g., cardiovascular reactivity and pupil dilation) have been linked to the provision of monetary incentive (Koelewijn et al., 2018, 2021; Richter, 2016). These physiological markers are also thought to correlate with subjective measures of how much effort a listener is exerting while understanding the speech signal (Peelle, 2018). While effortful listening is often used in understanding speech perception by hearing-impaired populations, similar issues are likely at work when listening to speech from an unfamiliar accent (Van Engen and Peelle, 2014).

Interestingly, while some previous work has demonstrated that participants show increased physiological responses in conditions of monetary incentive, they do not demonstrate improvement in behavioral outcomes (Koelewijn et al., 2018, 2021).

The dichotomy between those previous results and those presented here is interesting given that in some ways one would expect similar performance across these projects. However, the work here differs from the previous work on a few dimensions that may be relevant for understanding these diverging results. First, the task in this study was to transcribe speech produced by an unfamiliar talker with an unfamiliar accent. While the repetition task in Koelewijn and colleague's work is similar, the speech-in-noise task used there is, in some ways, more challenging than the task used here. Further, we were interested in adaptation across the course of an experiment; therefore, our measures were slightly different than those used in the previous work. Future studies could consider why the performance here diverges from this previous work.

Still, it is likely that the improvement in performance we see here is driven by increased attention or effort during listening. Therefore, while no current models of accent adaptation can fully account for the modulation in performance as a function of incentive we observe here, models of listening effort that are designed for understanding speech perception from hearing-impaired populations do include modulation of motivation. For example, the Framework for Understanding Effortful Listening (FUEL) (Pichora-Fuller *et al.*, 2016) integrates issues of motivational intensity (Brehm and Self, 1989) with other cognitive factors known to impact listening effort. This demonstrates the need for more-sophisticated models of accent adaptation that include not only aspects of the target speech (Baese-Berk *et al.*, 2013; Xie and Myers, 2017) but also other factors known to influence listening effort, including motivation.

Finally, it is crucial to understand the real-world implications for these results. As a reviewer notes, "We can't go around paying people to motivate them while they're engaging in conversations." Indeed, it is crucial to determine how both external and internal motivations might modulate these results. In a series of planned experiments, we will investigate how occupational hierarchy (*i.e.*, someone being your superior *vs.* your subordinate at work) might impact these results and how places of employment could incentivize adaptation to unfamiliar accents. In other work, we are investigating how various types of training could help listeners adapt in educational settings. While the real-world implications are not straightforward, it is critical that future work addresses this area.

## Conclusion

We open a new avenue to investigate how listeners can improve their understanding of non-native speech. In doing so, we shift the burden of communication from being something that solely rests with the speaker toward a more-equitable sharing across the two parties engaged in communication. Our experimental results point clearly to an opportunity for improving native and non-native communication—we find immediate and sustained performance differentials induced by incentivizing listeners. Moreover, listeners learn faster in the presence of incentives, leaving unincentivized control subjects further behind over the length of the experiment. As the consequences of having a non-native accent likely extend far beyond communication, we anticipate the longer-run welfare benefits associated with improving listeners'

communication to be large—more-effective business decisions, richer personal and professional relationships, and improved attitudes toward non-native speakers.

**Acknowledgements.** This work was partially supported by a grant from the National Science Foundation to [removed for review] and a grant from the Undergraduate Research Opportunities Program to [removed for review].

**Competing interests.** The authors declare none.

## Notes

1 See Lazear (2000) for theoretical implications of switching from hourly wages to piece rates—this establishes the theoretical support for our prior that performance increases with the incentive we introduce to treated subjects. Lazear (2000) also tests the model's predictions against data, finding “extremely large” productivity effects of between 20 and 36% of output. Increases in productivity are also documented in Shearer (2004), for example, where estimates of the gain in productivity associated with workers being paid piece rates are on the order of 20%. Also, see Seiler (1984) and Haley (2003) for additional examples.

2 We did not have strong priors as to the effect size of providing incentive. In setting the piece rate at \$2, we had envisioned leaving subjects roughly equivalent in their expected payments across treatment and control groups. As there is no mechanism for income effects to feed back into the experiment—secured by subjects not receiving feedback on their performance, or their payments until after the completion of the experiment—we are not concerned that payments to treated subjects are slightly higher on average.

3 Speakers of the stimuli include ALL\_011\_F, ALL\_018\_F, ALL\_030\_F, ALL\_035\_M, ALL\_039\_M, and ALL\_043\_M from the ALLSTAR corpus. Further information about these speakers is available at: <https://speechbox.linguistics.northwestern.edu/#!/home>

4 This picks up the effect of any variation in difficulty, for example, but also any systematic difference in the number of words correctly identified due to the number of words, for example. (We will consider this source of heterogeneity separately—in Table 3.)

5 Again we control for any difference in average performance that is due to specific sentences in  $\delta_s$  – This is true throughout, regardless of how we model the order in which those sentences are experienced.

6 Specifically, we adopt as our preferred specification, with both sentence and sentence-order fixed effects (as in Table 2, Column 4).

7 Note that these specifications vary from the earlier specification—the pooled specification of Table 2, Column 4—as the Sentence fixed effect now fit average differences specifically for the treated group (Column 1) and control group (Column 2).

8 In both treated ( $p = .287$ ) and control ( $p = .907$ ) samples we reject fourth-order polynomials.

## References

- 2018 American Community Survey 5-Year Estimates, Table B01003; *American Community Survey*. U.S. Census Bureau.
- Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *Journal of the Acoustical Society of America*, *133*(3), EL174–EL180.
- Banks, B., Gowen, E., Munro, K. J., & Adank, P. (2015). Cognitive predictors of perceptual adaptation to accented speech. *The Journal of the Acoustical Society of America*, *137*(4), 2015–2024.
- Bent, T., Baese-Berk, M., Borrie, S. A., & McKee, M. (2016). Individual differences in the perception of regional, nonnative, and disordered speech varieties. *The Journal of the Acoustical Society of America*, *140*(5), 3775–3786.
- Borrie, S. A., Barrett, T. S., & Yoho, S. E. (2019). Autoscore: An open-source automated tool for scoring listener perception of speech. *Journal of Acoustical Society of America*, *145*, 392–399. <https://doi.org/10.1121/1.5087276>
- Bradlow, A. R. & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729.



- Bradlow, A. R., Kim, M., & Blasingame, M.** (2017). Language-independent talker-specificity in first-language and second-language speech production by bilingual talkers: L1 speaking rate predicts L2 speaking rate. *The Journal of the Acoustical Society of America*, **141**(2), 886–899.
- Brehm, J. W. & Self, E. A.** (1989). The intensity of motivation. *Annual Review of Psychology*, **40**(1), 109–131.
- Brutt-Griffler, J.** (2005). Globalisation and Applied Linguistics: Post-imperial questions of identity and the construction of applied linguistics discourse. *International Journal of Applied Linguistics*, **15**(1), 113–115.
- Carlson, H. K. & McHenry, M. A.** (2006). Effect of accent and dialect on employability. *Journal of Employment Counseling*, **43**(2), 70–83.
- Cherry, E. C.** (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, **25**(5), 975–979.
- Clarke, C. M. & Garrett, M. F.** (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, **116**(6), 3647–3658.
- Cowen, T. & Tabarrok, A.** (2018). *Modern principles of economics*, 4 ed. MacMillan.
- Gluszak, A. & Dovidio, J. F.** (2010). The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review*, **14**(2), 214–237.
- Haley, M. R.** (2003). The response of worker effort to piece rates: evidence from the midwest logging industry. *The Journal of Human Resources*, **38**(4), 881–890.
- Huang, L., Frideger, M. & Pearce, J. L.** (2013). Political skill: Explaining the effects of nonnative accent on managerial hiring and entrepreneurial investment decisions. *Journal of Applied Psychology*, **98**(6), 1005.
- Kachru, B. B.** (1986). *The alchemy of English: The spread, functions, and models of non-native Englishes*. University of Illinois Press.
- Kinzler, K. D., Corriveau, K. H. & Harris, P. L.** (2011). Children's selective trust in native-accented speakers. *Developmental Science*, **14**(1), 106–111.
- Koelwijn, T., Zekveld, A. A., Lunner, T., & Kramer, S. E.** (2018). The effect of reward on listening effort as reflected by the pupil dilation response. *Hearing Research*, **367**, 106–112.
- Koelwijn, T., Zekveld, A. A., Lunner, T., & Kramer, S. E.** (2021). The effect of monetary reward on listening effort and sentence recognition. *Hearing Research*, **406**, 108255.
- Kutlu, E., Tiv, M., Wulff, S., & Titone, D.** (2022a). The impact of race on speech perception and accentedness judgements in racially diverse and non-diverse groups. *Applied Linguistics*, **43**(5), 867–890.
- Kutlu, E., Tiv, M., Wulff, S., & Titone, D.** (2022b). Does race impact speech perception? An account of accented speech in two different multilingual locales. *Cognitive Research: Principles and Implications*, **7**(1), 1–16.
- Lazear, E. P.** (2000). Performance pay and productivity. *American Economic Review*, **90**(5), 1346–1361.
- Lev-Ari, S.** (2015). Comprehending non-native speakers: Theory and evidence for adjustment in manner of processing. *Frontiers in Psychology*, **5**, 1546.
- Lev-Ari, S. & Keysar, B.** (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, **46**(6), 1093–1096.
- McGowan, K. B.** (2015). Social expectation improves speech perception in noise. *Language and Speech*, **58**(4), 502–521.
- McLaughlin, D. J., Baese-Berk, M. M., Bent, T., Borrie, S. A., & Engen, K. J. V.** (2018). Coping with adversity: Individual differences in the perception of noisy and accented speech. *Attention, Perception, & Psychophysics*, **80**(6), 1559–1570.
- Mirkovic, B., Debener, S., Schmidt, J., Jaeger, M., & Neher, T.** (2019). Effects of directional sound processing and listener's motivation on EEG responses to continuous noisy speech: Do normal-hearing and aided hearing-impaired listeners differ? *Hearing Research*, **377**, 260–270.
- Munro, M. J. & Derwing, T. M.** (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, **38**(3), 289–306.
- Nilsson, M., Soli, S. D., & Sullivan, J. A.** (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, **95**(2), 1085–1099.
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B.** (1994). Speech perception as a talker-contingent process. *Psychological Science*, **5**(1), 42–46.
- Peelle, J. E.** (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, **39**(2), 204–214.

- Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, 162 (1–2), 8–13.
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., et al. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, 37, 5S–27S.
- Richter, M. (2016). The moderating effect of success importance on the relationship between listening demand and listening effort. *Ear and Hearing*, 37, 111S–117S.
- Rogerson-Revell, P. (2007). Using English for international business: A European case study. *English for Specific Purposes*, 26(1), 103–120.
- Rönnerberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., . . . , & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, 7, 31.
- Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, 33(4), 511–531.
- Seiler, E. (1984). Piece rate vs. time rate: The effect of incentives on earnings. *Review of Economics and Statistics*, 66(3), 363–376.
- Shearer, B. (2004). Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment. *The Review of Economic Studies*, 71(2), 513–534.
- Sheppard, B. E., Elliott, N. C., & Baese-Berk, M. M. (2017). Comprehensibility and intelligibility of international student speech: Comparing perceptions of university EAP instructors and content faculty. *Journal of English for Academic Purposes*, 26, 42–51.
- Sidaras, S. K., Alexander, J. E., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, 125(5), 3306–3316.
- The Condition of Education.** (2020). *English Language Learners in Public Schools*. U.S. Department of Education; Institute of Education Sciences, National Center for Education Statistics.
- Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, 8, 577.
- Van Wijngaarden, S. J. (2001). Intelligibility of native and non-native Dutch speech. *Speech Communication*, 35(1–2), 103–113.
- Vaughn, C. R. (2019). Expectations about the source of a speaker's accent affect accent adaptation. *The Journal of the Acoustical Society of America*, 145(5), 3218–3232.
- Xie, X. & Myers, E. B. (2017). Learning a talker or learning an accent: Acoustic similarity constrains generalization of foreign accent adaptation to new talkers. *Journal of Memory and Language*, 97, 30–46.

---

**Cite this article:** Afghani, C., Baese-Berk, M. M., & Waddell, G. R. (2023). Performance pay and non-native language comprehension: Can we learn to understand better when we're paid to listen? *Applied Psycholinguistics* 44, 593–609. <https://doi.org/10.1017/S0142716423000322>