# 音声学と音声分析の能動学習を
# オンライン学習が容易にする北海道大学の実践事例

河合　剛

北海道大学外国語教育センター

goh@kawai.com http://goh.kawai.com/

# Case study of how online learning facilitates active learning of phonetics and speech analysis at Hokkaido University

Goh Kawai

Hokkaido University, Center for Language Learning

**Abstract**

Hokkaido University offers a 1-semester blended learning course on phonetics and speech analysis to advanced undergraduate and beginning graduate students. The online and face-to-face components of the course correspond to individual and group study. Before class, students learn online, and can say, record, and analyze speech at places and times of their choosing, instead of spending long hours at the laboratory together. During class, students present and discuss their lab work in person. Our course is an example of how online learning reduces the burden of classroom scheduling, and by monitoring assignments ensures that students prepare for class.

## 1 Introduction

Laboratory courses (i.e., courses involving experiments and analyses) require considerable hands-on time. In many cases the equipment is located only at the lab, forcing students and professors to work there often late into the night. In some cases, the equipment is more accessible. For instance, a laptop computer may suffice.

This paper is a case study of a take-home lab course in phonetics and speech analysis, where the lab component is taken as an online course so that students can say, record, and analyze speech at places and times of their choosing, instead of spending long hours at the laboratory together. The instructor is freed from tending to questions as they arise in the lab. Classroom time can be dedicated to presenting results and discussing analyses.

The keys to success are to (a) provide step by step instructions, and (b) assure that each student is prepared for class by rewarding high quality assignments.

The remainder of this paper describes the course offering, goals, and assignments of the course (section 2), the online, autonomous learning component that is delivered via an LMS (learner management system) (section 3), and the collaborative, active learning component that takes place face to face in the classroom (section 4). I conclude with my recommendations (section 5).

## 2 Course offering, goals, and assignments

Hokkaido University offers a 1-semester blended learning course on phonetics and speech analysis to advanced undergraduate and beginning graduate students. The course is taught wholly in English language, and is considered a CLIL (content language integrated learning) and STEAM (science technology arts engineering and mathematics) course if taken to meet the student's English language requirement.

The course offering describes the course as follows:

> This is a hands-on course where students acquire the technical skills for using computers to analyze spoken language.
>
> During class, you will use your computers to give presentations in English. Before and after class, you will do assignments using the Glexa system [note: the name of our LMS] and speech analysis software. You may ask for help in English or Japanese, in person or via email.
>
> This course is suited for students who desire to (a) use technical English, (b) present using computers, and (c) analyze time series.
>
> Required general skills: (a) Students need knowledge of technical English obtained, for instance, through Hokudai's English II (English Online) "technical writing" learning tasks. Reading assignments, software manuals, presentations, questions and answers, CALL assignments, and explanations on how to learn are in the English language. You will read, write, and speak technical English. (b) Students must bring their own computer to class, and present their assignments using projectors that I will provide. MacOS, Linux, and Windows are acceptable. Smart phones and tablets are not acceptable, because they do not run the software we will use in this course.
>
> Preferred skills: Knowledge of linguistics, statistics, experiment design, Fourier transforms, and spectral analyses are advantageous but not essential. We will

explain in class what you need to know.

Unimportant skills: Speaking or writing English language accurately is not important. What you want to say is more important than how you say it. No matter how many mistakes you make in words, phrases, or syntax, as long as you have an interesting message you will succeed. By the end of this course, you will learn how to communicate in English.

The course offering states that, after completing this course, students will be able to do the following (this is a partial list):

(1) Explain basic characteristics of spoken language (e.g., the differences between read speech and spontaneous speech; features of close-talking microphones; features of telephone-bandwidth speech).
(2) Display and interpret narrow-band and wide-band spectrograms, spectra, formants, and F0 tracks.
(3) Record read speech and spontaneous speech.
(4) Label speech at the word and phone levels.
(5) Measure speech rate using various methods (e.g., the number of phones, syllables, or words per unit time considering filled and/or unfilled pauses).

The course offering describes assignments:

Each week, you will receive assignments. Examples of assignments include installing software, reading instructions, collecting speech, analyzing waveforms, and preparing presentations.

Assignments are structured incrementally, and require substantial hands-on effort. You need to work on your own computer, because you will bring it to class to give your presentation.

Students are encouraged to work together outside of class to complete assignments. Learning together helps you understand why and how you do assignments. Working with friends prepares you for working in the real world, because most scientists and engineers work in teams.

You will use the freely available, excellent software package http://www.praat.org/.

Class meets for 90 minutes, 1 time per week, for 15 weeks. The classroom has multiple projectors for practicing different presentations simultaneously.

## 3 Online, autonomous learning component of the course

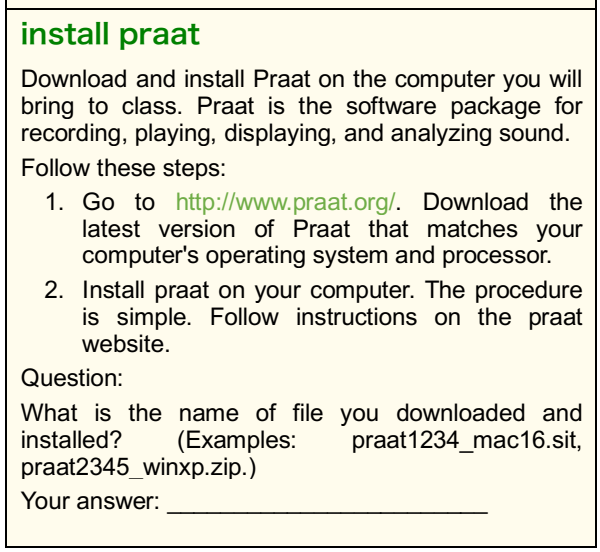
Autonomous learning requires scaffolding. Our class has 2 autonomous learning tasks: (1) say and analyze utterances, and (2) document findings by combining facts and logic. Both are reported by students in class. Task 1 is specific to this course, while task 2 is suited for a more general purpose. I will explain 1 first, followed by task 2.

### 3.1 Learning to say and analyze speech

We use Praat (http://www.praat.org/), which is a free and excellent software package for recording, playing, editing, analyzing, and labeling speech. Praat runs on Linux, MacOS, and Windows.

Showing how to use Praat is straightforward. We show step-by-step instructions using written explanations and screenshots (Figures 1, 2, 3). Although preparing the material is somewhat tedious, the invested time pays off because the instructor no longer needs to explain each step.

**Figure 1. Written instructions for installing software page.** This step is part of the assignment due by the 2nd class session. We indirectly monitor progress by asking students to write the name of the file they downloaded. (The filename changes often.)



install praat

Download and install Praat on the computer you will bring to class. Praat is the software package for recording, playing, displaying, and analyzing sound.

Follow these steps:

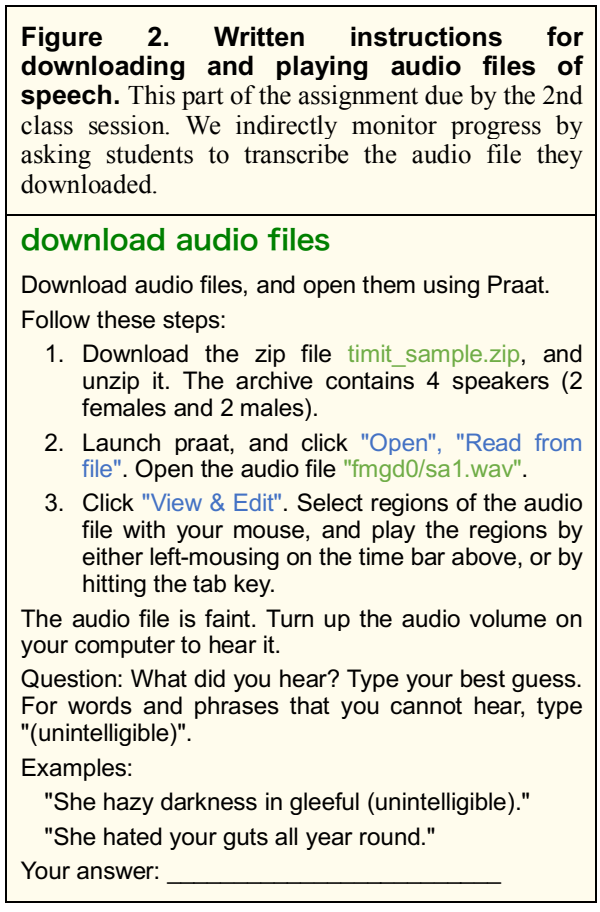
1. Go to http://www.praat.org/. Download the latest version of Praat that matches your computer's operating system and processor.
2. Install praat on your computer. The procedure is simple. Follow instructions on the praat website.

Question:

What is the name of file you downloaded and installed? (Examples: praat1234_mac16.sit, praat2345_winxp.zip.)

Your answer: _____

**Figure 2. Written instructions for downloading and playing audio files of speech.** This part of the assignment due by the 2nd class session. We indirectly monitor progress by asking students to transcribe the audio file they downloaded.



download audio files

Download audio files, and open them using Praat.

Follow these steps:

1. Download the zip file timit_sample.zip, and unzip it. The archive contains 4 speakers (2 females and 2 males).
2. Launch praat, and click "Open", "Read from file". Open the audio file "fmgd0/sa1.wav".
3. Click "View & Edit". Select regions of the audio file with your mouse, and play the regions by either left-mousing on the time bar above, or by hitting the tab key.

The audio file is faint. Turn up the audio volume on your computer to hear it.

Question: What did you hear? Type your best guess. For words and phrases that you cannot hear, type "(unintelligible)".

Examples:

"She hazy darkness in gleeful (unintelligible)."

"She hated your guts all year round."

Your answer: _____

Next week in class, you will open more files, display them, and play them.

To prepare for class, follow these steps:

1. Using the same file you opened above, click "Modify", "Scale to peak", new absolute peak = 0.99, "ok".
2. Click "View & Edit".
3. Play the file. Notice that the audio is much louder. Can you explain why?
4. The zip file that you downloaded contains audio files for 4 people. Open all audio files named "sa1.wav" and "sa2.wav". Listen to them. What do you notice?

We will ask each other questions and answer them in class.

**Figure 3. Partial written instructions and screenshot for labeling speech at the word and phone levels.** This step is part of the assignment due by the 4th class session. We monitor progress by asking students to upload files they analyzed.

We want our students to ask questions in class. Creating questions is part of the assignment. For each analysis, we ask students to write questions. In our assignments, we do not expect students to deduce all the answers (our assignments prepare students for the next class, where the answers are obtained), but we do expect students to identify aspects that puzzle them.

## Label speech at the word and phone levels

[... several steps omitted ...]

Last week, you learned the following rules:

1 Each and every word in the utterance must appear in the "word" tier.

2 Words may not overlap.

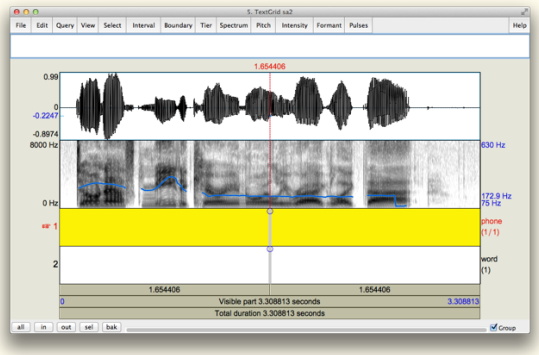
3 Every part of the utterance must be labeled. No gaps, no empty spaces.

This week, add these rules:

4 A word boundary must also be a phone boundary. That is, where there is a word boundary, there must be also a phone boundary.

5 Phones may not overlap.

6 Multiple words or phones that cannot be separated may be labeled as one unit. For instance, if you cannot divide the 2 words "oh wow" into "oh" and "wow", you may label them as "oh wow" without a word boundary. Same rule for phones.



[... several steps omitted ...]

Label the utterance "sa2.wav". Choose a talker other than the talker you chose last week.

The utterance contains the words "Don't ask me to carry an oily rag like that.".

When you finish labeling the utterance, upload your "Collection" file [note: a bundle of files showing what was analyzed how] to Glexa. Click on the file upload button below.

Bring your file to class.

Bring your computer also, if you wish to practice using it. You may use my computer if you wish. I recommend bringing your computer, because it gives you practice in presentations. The most common cause of anxiety and failure in presentations is connecting your computer to the projector and speakers.

When you label your speech file, you will be puzzled and mystified. Bring your questions to class. Asking questions is the most important skill for scientists.

Type your questions in the box below. You may write in English, Japanese, or Dutch (these are the languages I understand). _____

Questions? Problems? Contact me via GlexaMail [note: LMS's email tool] link towards to upper right corner of your browser.

### 3.2 Learning to document findings

Students document their findings by combining facts and logic. Students show how they analyzed speech by demonstrating speech analysis software in class.

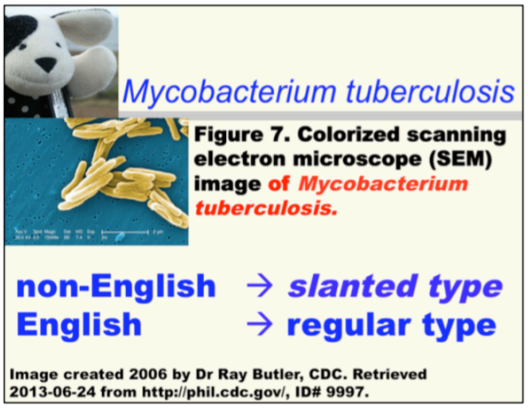
Although students already know how to use Microsoft PowerPoint or Apple Keynote, students do not create many slides, because the speech analysis software provides a common platform for sharing ideas, and the explanations, questions, and answers are mostly oral. The oral presentation (and scripts, if students want to write them) do require knowledge of technical writing.

Technical writing is introduced to all Hokudai freshmen via a required online course in English language (Figure 4). In subsequent semesters, we repeat training of technical writing so that upperclassmen warn freshmen that technical writing is used throughout their undergraduate career. Our teaching technique is based on our observation that students are not necessarily keen on acquiring skills. They want better grades. Advice from peers encourages study that indirectly achieves our objectives -- namely, to enable students to create and give technical presentations by themselves.

**Figure 4. Screenshot of an online task for learning technical writing.** This is part of a separate course required of all freshmen during their 1st semester at our school.

The rules for technical writing in English language are rigid and clearly defined. The standards for good writing are so widely known that American college school graduates can judge the quality of writing, even if they cannot produce good writing themselves.

Our students prefer the restrictive, prescriptive, unambiguous nature of technical writing.

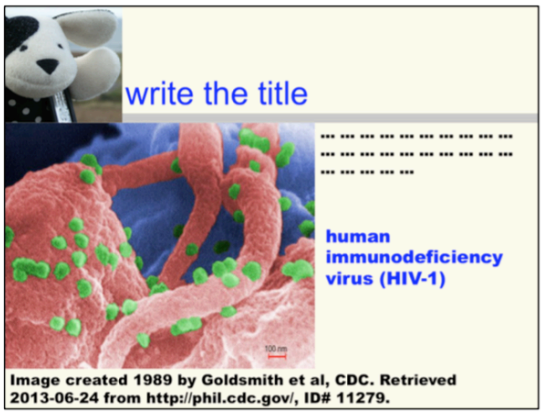



What does the image show?

Because we're looking at *Mycobacterium tuberculosis*, our title becomes "Figure 7. Colorized scanning electron microscope (SEM) image of *Mycobacterium tuberculosis*".

We write scientific names in slanted type. Most scientific names are of Latin or Greek origin.

We write English words in regular, non-slanted type.




Write the title for the image.

# 4 Collaborative, active learning component of the course

## 4.1 What students do

The class is flipped. Students prepare for class by completing online assignments that students show or say in class. Students submit assignments on our LMS instead of writing on their textbooks (the latter is common practice among language educators in Japan). Online submission allows students to record their utterances and upload their assignments, so that instructors can comment before class (instead of collecting handwritten material during class to be returned at the next class session).

Each class begins with presentations by individual students (Figure 5). We emphasize that asking questions and answering them is a crucial skill. Students are encouraged (but not required) to prepare slides, especially slides designed specifically for answering questions (that is, students are taught not to not show all slides during their talk). At least half of the time given to each presenter is allocated for Q&A. Students are graded based on their presentations they give, the questions they ask (when they are members of the audience), and the answers they give. We reward talks that receive many questions.

**Figure 5. Photograph of student presenting.** The photograph has been blurred to conceal the identities of the students.

Students are encouraged to bring their own laptop computers to class, so that students can practice giving technical presentations later in their careers. Students connect their laptops to our projectors and loudspeakers. Classmates are encouraged to ask questions at any time.

Each class introduces English language expressions used in reporting technical procedures and discoveries. Each student reports in each class. Because each week the expressions and procedures were repeated (with new ones being added), students memorized the phrases and process. Our students are not skilled in English language (typically CEFR A2 or B1), but become adept at reporting their laboratory findings in class. By the end of the semester, students have repeating tasks so often that they have become a habit.

When the class size is large, we provide a group learning task, because scientists and engineers often work in groups, and because group interactions encourage speaking English and drawing diagrams (Figure 6).

Groups are randomly assigned each class period. Fixed groups are unpopular (students complain of freeloaders) and unmanageable (absent or late students create imbalance across groups, particularly in classes where the same students tend to be absent or late in every session).

**Figure 6. Photograph of classroom in group work layout.** This photograph is from a class session with 16 students. The photograph has been blurred to conceal the identities of the students.

Sets of 4 desks form inward-facing rectangles. This layout facilitates eye contact, talking, writing, and drawing. The rectangular layout forces students to turn around when watching videos or demonstrations. Leaving sufficient space around the desks allows students to rotate their chairs. We do not move desks during class, although some instructors of active learning courses assemble and dissolve groups by placing desks together or apart.

### 4.2 What instructors do

Instructors demonstrate and motivate during class. Face-to-face class time is for hand-on experimentation.

A prime example is the reed, tube, and plunger combination that produces various vowels [1][2][3][4][5][6]. This seemingly simple apparatus cannot be purchased in quantity because it is rather expensive. Students take turns trying it in class (Figure 7).

We also use various tools for teaching phonetics [7], although the best method by far is a trained human being demonstrating while students imitate.

## 5 Conclusion and recommendations

### 5.1 Summary

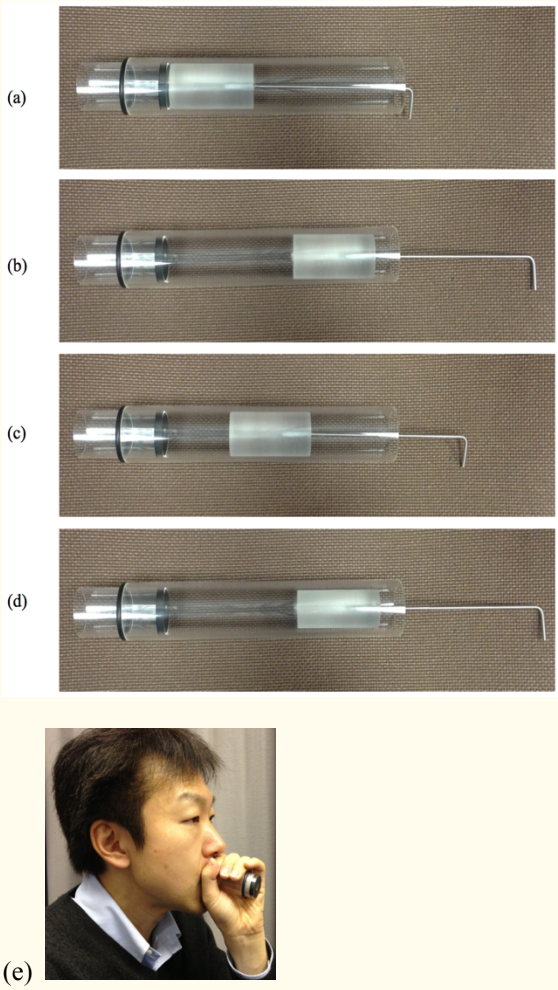
The key to training students is to give all and expect all [8][9]. Giving is good. Babying is bad.

The key to flipping classes by leveraging online learning is finding repetitive tasks that students can do alone [10][12]. Repeat tasks until they become a habit.

The key to activating classes in flipped classes is to give students authority and responsibility in directing the class session [13]. Instructors are coaches. Stay on the sideline.

The key to training ourselves as teachers is to recognize and grow our capabilities [14]. Be a good student.

**Figure 7. Photograph of (a)-(d) reed, tube, and plunger for producing vowel-like sounds, and (e) the inventor playing the reed.** Takayuki Arai (Sophia University) is committed to teaching acoustics. His apparatus visualizes the configuration of the vocal tract with regards to vowel-like sounds.




### 5.2 Details

Active learning is not new. In medieval times when education was accorded to the elite minority, tutors would teach a small number of students at a time. Questions from both teacher and student were more frequent, while lectures were less. Passive learning became mainstream when education became available to the masses. Lectures increased; interaction decreased. School time was divided more rigidly into various subjects -- 2 class periods a week for mathematics, 4 for language, and so forth -- because teachers started to specialize in subject matter. As the daily, weekly, and yearly class schedules became inflexible, teachers became unable to spend much time (or, more importantly, elastic lengths of time) with each individual student. Teachers were faced with a dilemma

-- should we target the top students (who arguably have the most to contribute to society after finishing school), the middle students (who are the most numerous), or the bottom students (who deserve quality education just as much as anybody)?

Blended, flipped learning provides the means to bring back active learning to the classroom. The passive or independent learning tasks can take place outside of class at times and places of the student's choosing. The active or interactive learning tasks can take place inside of class. Students happily complete homework as long as they are graded. A typical student is less concerned of learning the subject matter than earning high grades. The wily instructor guides students towards studying by offering many tiny rewards to many tiny tasks. Yes, grading many tasks accurately and fairly is a lot of work. No, online learning or active learning do not save time or money. Instead, they help students achieve more by spending more time and thought. Intentional learning rewards us.

We have forgotten (or worse, never experienced) how to learn actively because we are inundated with passive learning. Few professors at my school feel comfortable in a classroom without desks or chairs. But elementary school teachers are perfectly capable of teaching in the schoolyard. We need to relearn active learning techniques.

One way to become an effective teacher is to become an effective student. Every several years I start learning something new. Some learning takes place behind a desk. Much takes place on the field or in the practice room. I learn new teaching techniques. I am reminded what it feels like to be a helpless beginner (I am a klutz with physical exercise and performing arts). I ponder how to empathize with students.

One way to improve the student-to-instructor ratio (and perhaps to decrease the workload of grading homework) is by hiring or giving course credit to upperclassmen as mentors and courseware developers. We solidify our knowledge by imparting it to others. Students with high English language proficiency who are exempted from English language courses are prime candidates for coaching underclassmen. Freshmen receive quality training, and sophomores enhance their mentoring skills. Our courses will benefit from courseware designed by students for students. Learning is a creative democracy. The best ideas sometimes come from people closest to the learners.

## 6 References


[1] Takayuki Arai "Mechanical models of the human vocal tract" Acoustics Today, 9(4)25-30, 2013

[2] Takayuki Arai "Education in acoustics and speech science using vocal-tract models" Journal of the Acoustical Society of America, 131(3)2444-2454, 2012

[3] Takayuki Arai "Simple physical models of the vocal tract for education in speech science" Proceedings of INTERSPEECH-2009, 756-759, Brighton, 2009

[4] Takayuki Arai "Education system in acoustics of speech production using physical models of the human vocal tract" Acoustical Science and Technology, 28(3)190-201, 2007

[5] Takayuki Arai "Sliding three-tube model as a simple educational tool for vowel production" Acoustical Science and Technology, 27(6)384-388, 2006

[6] Takayuki Arai "The replication of Chiba and Kajiyama's mechanical models of the human vocal cavity" Journal of the Phonetic Society of Japan, 5(2)31-38, 2001

[7] Jacqueline Vaissiere "New tools for teaching phonetics" Proceedings of International Conference on Phonetic Sciences (ICPhS), Barcelona, 2003

[8] Goh Kawai "Description of an online learning course for English language at Hokkaido University" Eurocall Conference (Eurocall 2017), Southampton, UK (2017-08-24)

[9] Goh Kawai "Proficiency of college freshmen can change substantially within an 8-week period" Japan Association of Language Teachers Special Interest Group in Computer Aided Language Learning Conference (JALT-CALL 2017), Matsuyama, Japan (2017-06-17)

[10] Akio Ohnishi and Goh Kawai "Pre-matriculate mobile-assisted language learning for studying before entering college" Teaching English to Speakers of Other Languages Conference (TESOL 2017), Seattle, Washington, USA (2017-03-23)

[11] Naomi Suzaki and Goh Kawai "Encouraging blended learning via a group singalong activity for learning segmental and prosodic pronunciation features" Teaching English to Speakers of Other Languages Conference (TESOL 2017), Seattle, Washington, USA (2017-03-22)

[12] Ivy Chu-hui Lin and Goh Kawai "Identifying and activating receptive vocabulary by an online vocabulary survey and an online writing task" in Salomi Papadima-Sophocleous, Linda Bradley, and Sylvie Thouësny, editors "CALL communities and culture – short papers from EUROCALL 2016", 271-276, article PDF https://research-publishing.net/publication/chapters/ 978-1-908416-44-5/574.pdf (2016-12-18)

[13] Akio Ohnishi and Goh Kawai "Active learning of language using hand-written comments" Association for the Advancement of Information and Communication Technology at Universities Conference (AXIES 2015), Nagoya, Japan (2015-12-04)

[14] Noriaki Katagiri and Goh Kawai "Vocabulary size limits what middle school non-native Instructors can say in all-English classes of English language" Proceedings of the American Association for Applied Linguistics Conference (AAAL 2015), Toronto, Ontario, Canada (2015-03-21)


## 7 About the author


Goh Kawai designs and uses online learning systems for autonomous or blended learning. Goh has a BA in linguistics, an MA in educational technology, and a PhD in information and communication engineering.