Getting the best out of Speech Technologies for pronunciation teaching

Beata Walesiak & William Gottardi

Beata Walesiak has worked as a teacher and teacher trainer specialising in pronunciation and technology-supported teaching. Currently, she is working as linguist for a mobile pronunciation app (ELSA) and continues her work via unpolish.pl.

William Gottardi is an MA candidate in English Linguistics at the Federal University of Santa Catarina (Brazil) and holds a BA in Software Analysis and Development and a BA in Portuguese and English Language Teaching.

Beata Walesiak & William Gottardi, 2022

Today, speech technologies!

Some research on their effectiveness

Practical application: speech technologies in web browsers & mobile apps

Opportunities behind them & their limitations

Our target learners: Teenagers & adults

Speech technologies

Automatic Speech Recognition (ASR)

Transcribes speech
based on **oral input**, also
known as STT (Speech-to-Text)

(Moussalli & Cardoso, 2020)

Text-to-Speech (TTS)

Automatically generates
synthesized speech
from units of **written text**
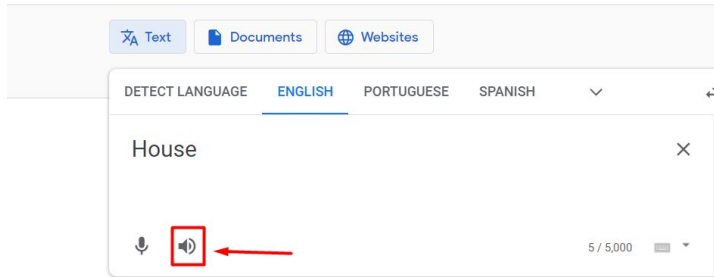
(Liakin et al., 2017)

ASR

web browsers

voice typing

smart assistants

mobile apps & conversational agents
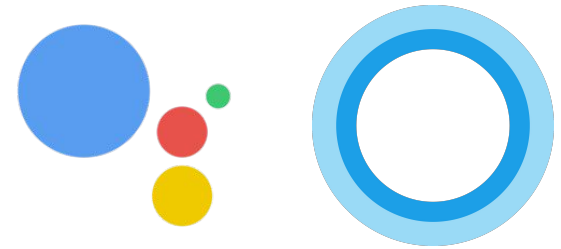
TTS

online translators

smart assistants

immersive readers

Research on ASR-supported pronunciation training

- Increased learner autonomy in & out of the classroom
  (Levis & Suvorov, 2013; Liakin et al., 2017; McCrocklin, 2014, 2016; Bashori et al., 2020, 2021)
- More willingness to communicate & reduce foreign language anxiety
  (Inceoglu et al., 2020; Kim, 2006)
- Oral production improvement
  (Neri et al., 2006, 2008)
- Limitless opportunities to practice speech
  (Dizon & Tang, 2020; Walesiak, 2021)
- Facilitating the evaluation of pronunciation
  (Fouz-González, 2015)
- Positive influence on students' learning process
  (Mroz, 2018; Walesiak, 2021)

Research on TTS-supported pronunciation training

- Limitless oral input & useful for out-of-class assignments
(Cardoso et al., 2015)
- Efficient & personalized feedback
(Cardoso, 2018)
- Raising learners' awareness to specific features and forms
(Liakin et al., 2017; Gomes et al., 2018)
- Increased learner autonomy in their own phonological development
(Moon, 2012; Liakin et al., 2017)
- Reinforcing grapheme-phoneme relationships
(Handley, 2013)

Opportunities & practical application

ASR & TTS on the web

Methodologies for ASR & TTS

Communicative Framework for Teaching Pronunciation

(Celce-Murcia et al., 2010, p. 45)

1. Description and analysis

2. Listening discrimination

3. Controlled practice

4. Guided practice

5. Communicative practice

Focus on **intelligibility**

Methodologies for ASR & TTS

Criteria for CALL task appropriateness

Chapelle (2001, p. 55)

1. Language learning potential

2. Learner fit

3. Meaning focus

4. Authenticity

5. Positive impact

6. Practicality

ASR & TTS put together?

- Target-like **oral input** (TTS) + opportunities to practice **oral output** (ASR)

- **Complement** to classroom instruction and suitable for extra-class activities

- Potential for **individualized instruction**

- For the **learner**: interaction opportunities, feedback, more contact with the target language, increased motivation and interest

(Gottardi et al., 2022)

Activity: **Google Pronunciation Practice**

Resources:
- **Google word definition**
- **Google Pronunciation Practice**

Schedule	American pronunciation ▾
Sounds like **skeh · jool** 🔊	
⬜ Slow	🎤 Practice
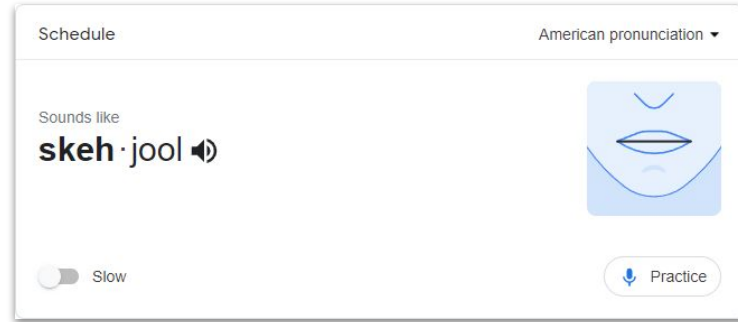
Tips:
- Say: "_____" definition (ex: casual)
- Type: "how to pronounce _____" (**complex words**)
- Teach expressions and phrases:
    - "rain on someone's parade" definition
    - "kick the bucket" definition
    - "go with the flow" definition

Activity: **Dictation of words, sentences & texts**

Resource: **speechnotes.co**

Tips:
- Allow the webpage to use your microphone.

Activity: **Google review**

Resources:
- **Tophonetics.com**
- **Google Docs**
  - **Shared document**

Tips:
- 4 skills integration
- Check spelling, grammar and clarity

Activity: **Shadow reading**

Resources:
- **Microsoft Edge (Immersive Reader)**
- **Authentic Material**
- **Google Translate (ASR)**

Tips:
- Authentic material
- Use different accents
- Look for **tutorials** on how to use the digital resources

Activity: **Minimal pairs practice**

Resource: **voicenotebook.com/prononce.php**

Tips:
- Use the alternative box to check homophones.
- There is also a mobile app available.

ASR & TTS on mobile

Mobile apps & methodologies

Apps or their functionalities	ASR	TTS	Suitable for
Dictionary apps or translators with voice search	✓	✓	Description and analysis Listening dicrimination tasks Controlled & Guided practice
Default OS* apps, e.g. web browsers, voice recorders	✓	✓	Controlled & Guided practice
Apps designed specifically for language learning	✓	✓	Controlled & Guided practice Communicative practice (some)
In-app assessment features using AI**-generated feedback	✓	✓	Needs analysis Summative or formative assessment
Conversational agents / virtual voice assistants	✓	✓	Controlled & Guided practice Communicative practice*

*OS = Operating System

**AI = Artificial Intelligence

(Walesiak, in prep. after Rosell-Aguilar, 2017 and Celce-Murcia et al., 2010)

Activity: **App-based dictations**

Resource: **Google Translate app**

Tips:

- For **controlled practice** on /æ/, ask learners to dictate words/shorter phrases with the target feature and see how many times the app got it right, e.g. ***ankle in pain.***
- Collect the ASR output, e.g. with the help of the Mentimeter word cloud functionality.

- In the screenshot, out of 25 attempts made by 2 students & 1 teacher, 11 were successful.
- This also raises learners' awareness and awakens their curiosity **to test the app** more.

How did Google Translate understand you?

i'm going painting
i'm calling pain   ankle in pains
uncle in pain   emily peyton
ankle in pain
aamco in pain   england pain
ankle and pain
encoding pain
how colon pain   ankle pain

Beata Walesiak & William Gottardi, 2022
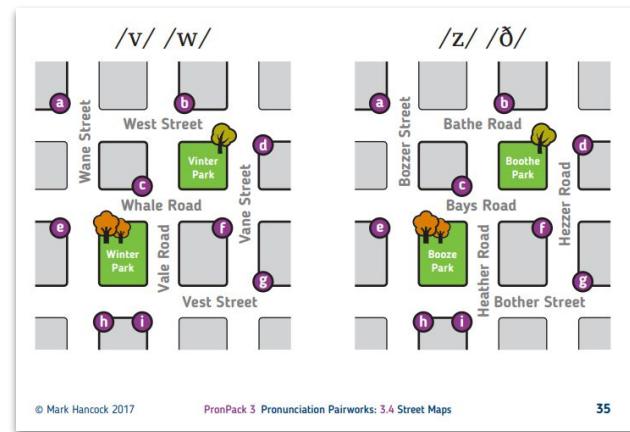
Activity: **App-based dictations**

Resource: Google Translate app

Tips:

- Adapt ready-made tasks to use them with the ASR.
- **Meaningful** gap-info tasks can be found in PronPack and used for pairwork, e.g. asking the way with the help of a street map (PronPack 3.4).

Shall I take the Bathe Road?
vs *Shall I take the Bays Road?*

Go along the Vale Road.
vs *Go along the Whale Road.*

/v/ /w/   /z/ /ð/

© Mark Hancock 2017   PronPack 3 Pronunciation Pairworks: 3.4 Street Maps   35

Activity: **App-based dictations**

**Resource:** ColorNote app

Tips:

- For longer dictations, you may want to use any **notepad app** that uses <u>keyboard</u> **voice input/typing,** e.g. ColorNote app.
- Ensure your learners use **the same target variety** for similar results**.**
- Choose a longer text and help your learners **prepare** to read it out, include:
  - inner voice, mouthing, mirroring & shadowing activities
- Ask your learners to dictate the text to the ASR (multiple attempts), then to work **in pairs** and **spot the similarities or differences against the original text**.
- Encourage **reflection**:
  - *How did the ASR work with both of us?*
  - *Is there something we could improve?*
  - *Is it us or the ASR? How do we know?*
  - *Which ASR is more reliable?*

Activity: **App-based dictations**
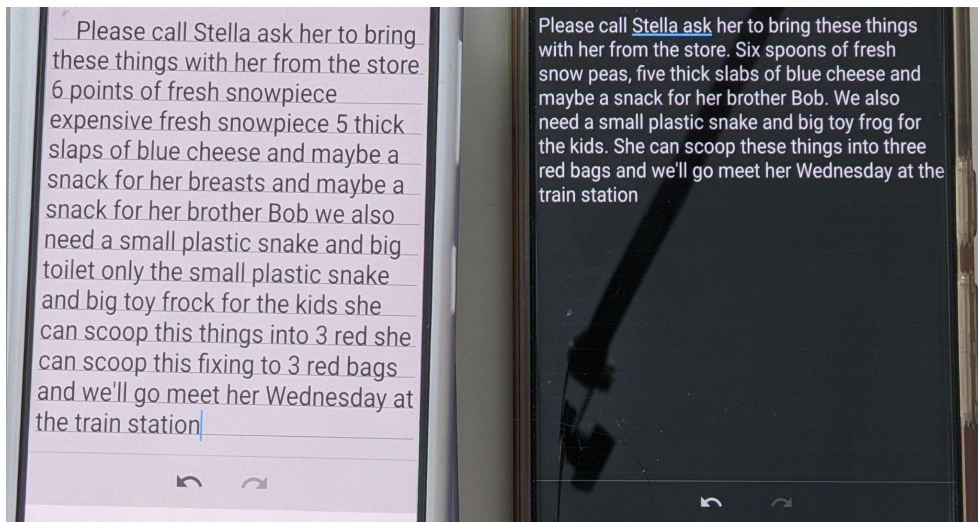
Samsung vs Google in action:
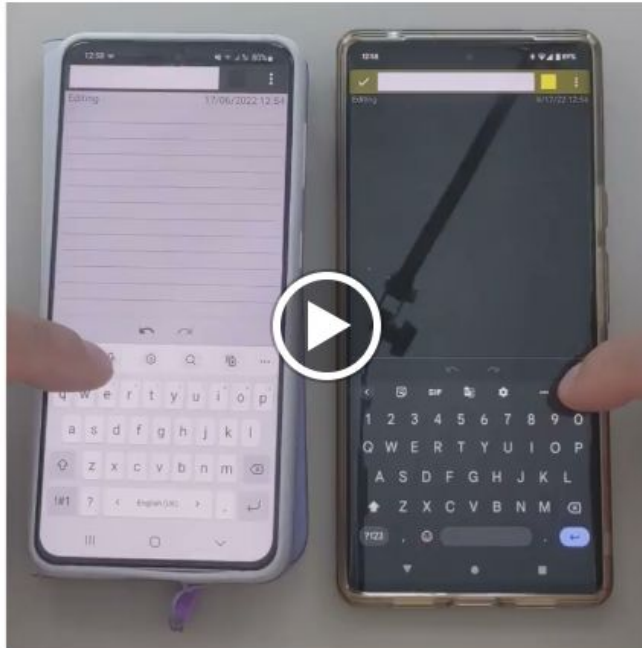
Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.

Stella (source: Speech Accent Archive)

Activity: **App-based dictations**

Samsung **vs** Google:

Please call Stella ask her to bring these things with her from the store 6 points of fresh snowpiece expensive fresh snowpiece 5 thick slaps of blue cheese and maybe a snack for her breasts and maybe a snack for her brother Bob we also need a small plastic snake and big toilet only the small plastic snake and big toy frock for the kids she can scoop this things into 3 red she can scoop this fixing to 3 red bags and we'll go meet her Wednesday at the train station

Please call Stella ask her to bring these things with her from the store. Six spoons of fresh snow peas, five thick slabs of blue cheese and maybe a snack for her brother Bob. We also need a small plastic snake and big toy frog for the kids. She can scoop these things into three red bags and we'll go meet her Wednesday at the train station

*Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and **a** big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.*

Stella (source: Speech Accent Archive)

Activity: **Needs analysis**

Resource: ELSA app

Tips:

- For needs analysis use ELSA's **Assessment Test** functionality.
- To find the full Assessment Test test go to: *Profile → Progress → Take the test*

- ELSA's AI prepares a detailed report on the learner's pronunciation as well as personalised tasks
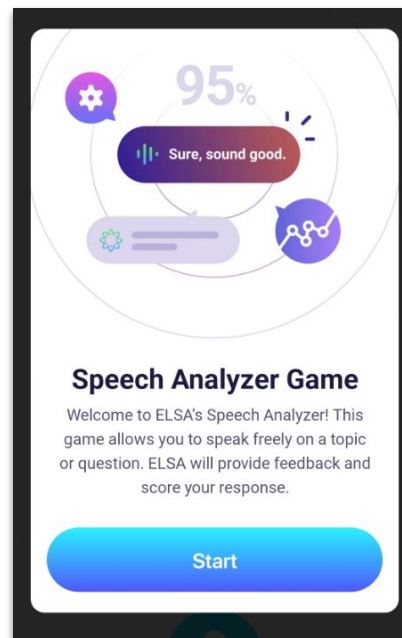- For now, General American is the target variety.
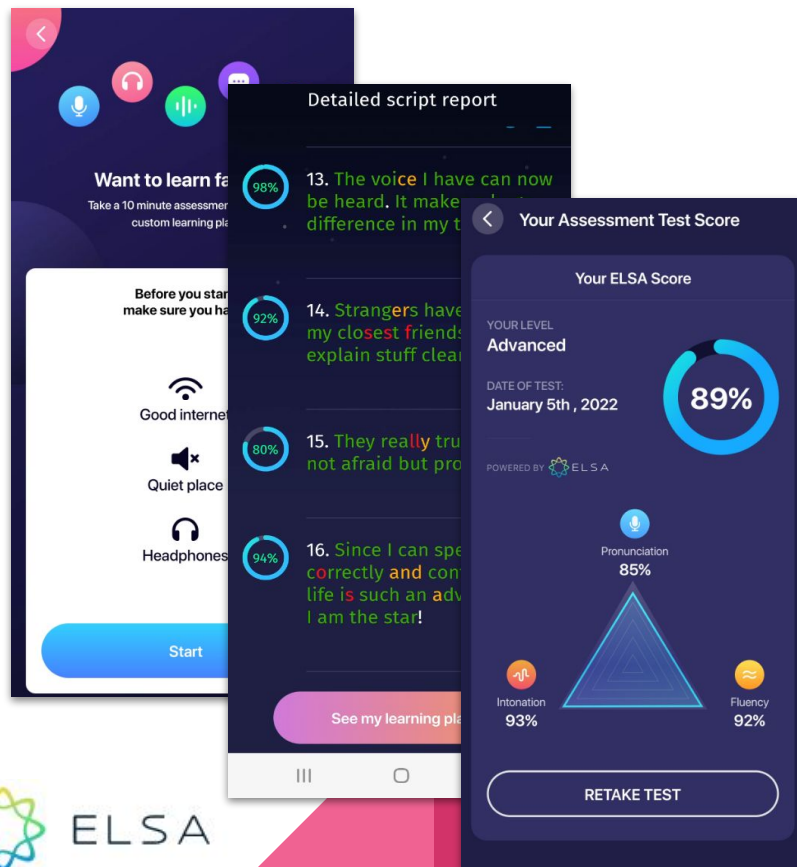
Activity: **One-way free practice**

Resource: ELSA app

Tips:

- Use **ELSA Speech Analyzer** for one-way communicative practice.
- To make your learners' practice relevant, remember that the questions include:
  - *Mock job interview*
  - *Daily English practice*
  - *Mock IELTS Speaking exam (parts 1-3)*

- The tasks enable the user to practice free speech and receive feedback on their delivery.
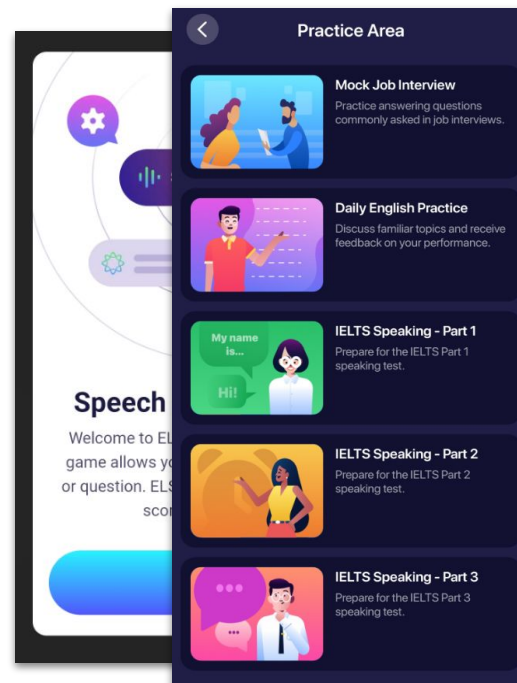

Speech Analyzer Game

Welcome to ELSA's Speech Analyzer! This game allows you to speak freely on a topic or question. ELSA will provide feedback and score your response.

Start

ELSA

Activity: **One-way free practice**

Resource: ELSA app

Tips:

- Use **ELSA Speech Analyzer** for one-way communicative practice.
- To make your learners' practice relevant, remember that the questions include:
  - *Mock job interview*
  - *Daily English practice*
  - *Mock IELTS Speaking exam (parts 1-3)*

- The tasks enable the user to practice free speech and receive feedback on their delivery.

Activity: **One-way free practice**

Resource: ELSA app

Tips:

- Use **ELSA Speech Analyzer** for one-way communicative practice.
- To make your learners' practice relevant, remember that the questions include:
  - *Mock job interview*
  - *Daily English practice*
  - *Mock IELTS Speaking exam (parts 1-3)*

- The tasks enable the user to practice free speech and receive feedback on their delivery.
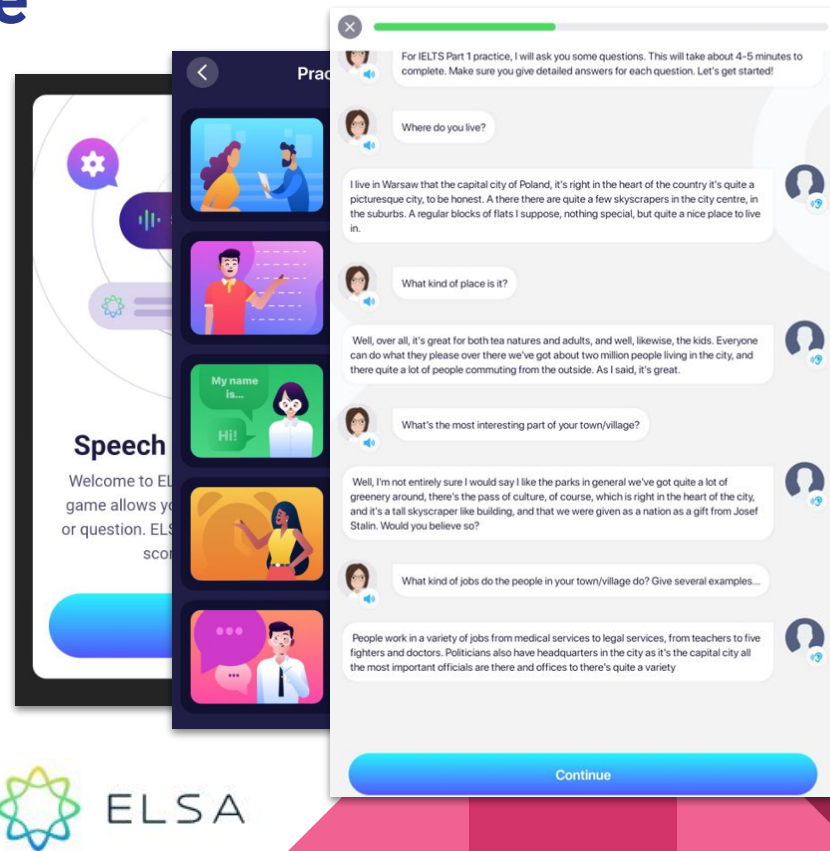
Activity: **Two-way 'conversation'**

Resource: Google Assistant

Tips:
- Start with simple vocabulary and fact finding
- Encourage your learners to work in-class or as home assignment to look up facts, names, locations, etc. (weather is a good start).
- For guided practice, encourage experimenting with the Google Assistant and prepare a few basic questions for lower-level students with the desired feature.
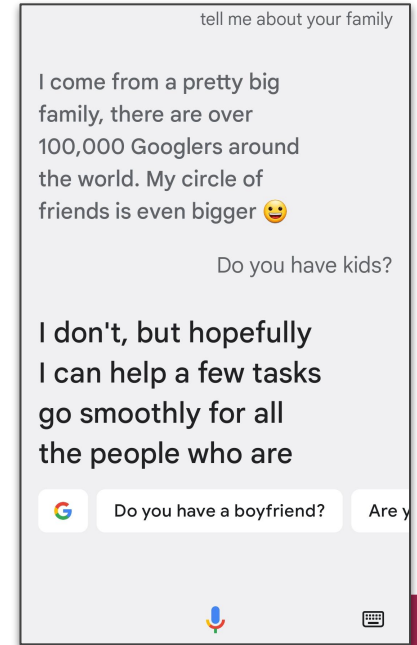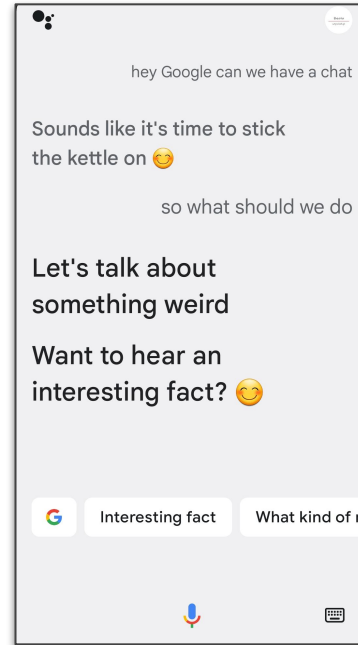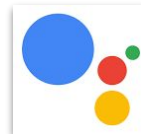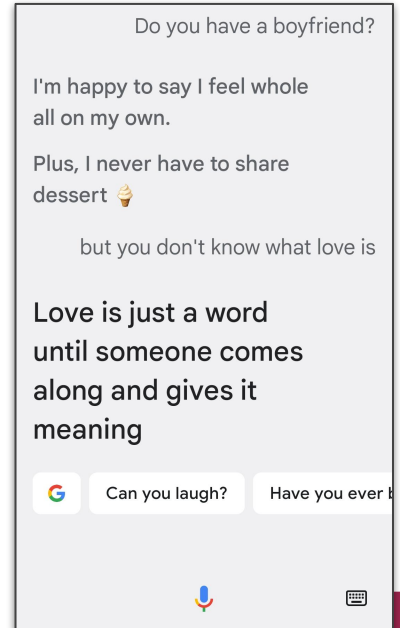
Activity: **Two-way 'conversation'**

Resource: Google Assistant

'We want you to be able to just talk to the Google Assistant like you would with another human.' ([Nino Tosca](), 2022)

Tips:
- Additional [Google Assistant games & features]() specify what commands to ask the GA to receive some feedback.
- **Potential pairwork**: Ask your learners to design a conversation they would have with the GA using the key sound/minimal pair/phrase/tone and then based on the mistakes they observe, ask them to redesign it.

Do you have a boyfriend?

I'm happy to say I feel whole all on my own.

Plus, I never have to share dessert 🍦

but you don't know what love is

Love is just a word until someone comes along and gives it meaning

| G | Can you laugh? | Have you ever |

Activity: **Two-way 'conversation'**

Resource: Google Assistant

Tips:

- Ask about a fact.
- Make a reference to an item in the answer.
- Ask some *Yes/No* or *Why/How* questions.
- Use the prompts that GA gives you.
- React to GA's answers.
- Come up with your own questions and see where it takes you!
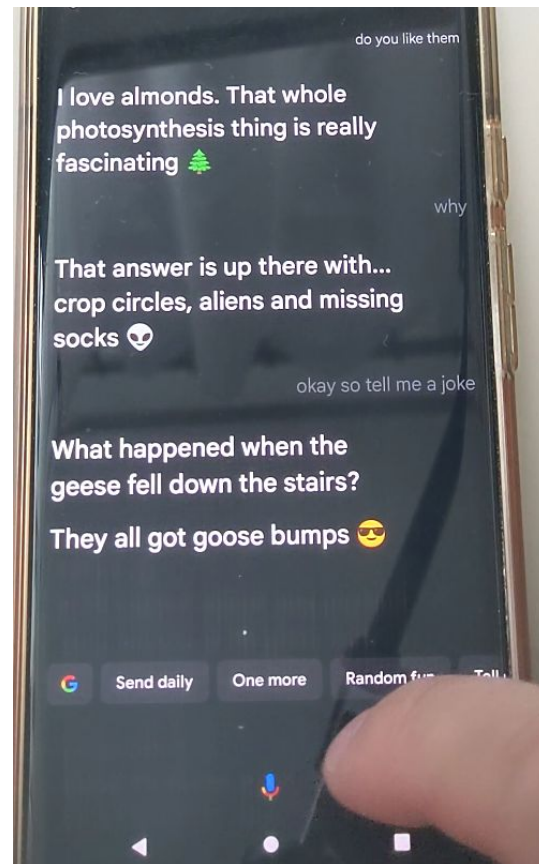
Activity: **Two-way 'conversation'**

Resource: Google Assistant

e.g.
Lesson topic: Healthy food
Aim: Practice falling & rising tones

\ *What is the most healthy food?*
\ *What are almonds?*
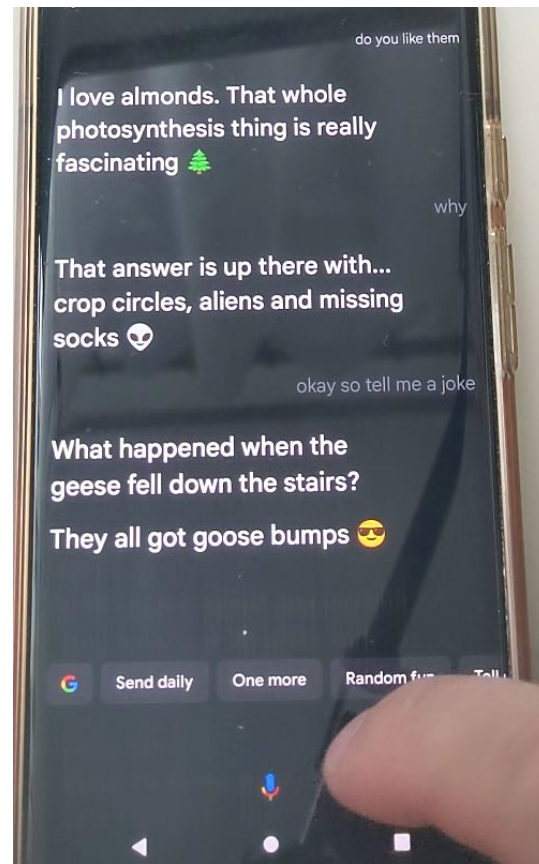/ *Do you like them?*
\ *Why?*
\ *OK, so tell me a joke.*
\ *You're funny.*
/ *Can you laugh?*
/ *Can you cry?*
\ *That sounds great!*

Limitations & pedagogical implications

ASR & TTS limitations

- **Using ASR & TTS you need to remember that**:
  - Use **predictable** vocabulary (avoid proper nouns, some combinations of words)
  - Some speaker characteristics **may affect quality** (e.g. child speech)
  - Paid access for longer texts/audio may be required
- **Not all TTS** may:
  - Offer accent variation
  - Offer natural speech rate* (a potential advantage for lower-level learners!)
- **ASR accuracy & WER:**
  - Google's accuracy with NNS speech ranged from 88.61% to 93.47% (McCrocklin et al., 2019)
  - **Word Error Rate (WER)** is below 10% for voice search & voice typing
    - Computational power & more data → lower WER
    - Demand for ASR & TTS → speech technology improvements

(Yu & Deng, 2015; Eksi & Yesilcinar, 2016; Jurafsky & Martin, 2021; Gottardi et al., 2022)

Pedagogical implications

'Educational technology is only as good as the **humans** behind it.'
(Revell-Rogerson, 2021)

- The presented resources can:

  - **promote learner autonomy** & focus on learners' specific **needs**

  - be used in **various learning contexts**

  - facilitate **integrating the pronunciation** component in language courses

  - support **some varieties** of English (test your variety beforehand)

Pedagogical implications

- Make sure that you try out and do the **app/tool content analysis** first!

- To achieve better results, ensure access to:
  - a good-quality **microphone**
  - **stable internet** access
  - **quiet** environment

- If something on the app doesn't work, **report the bug** to the app developer. This helps the people at the backend fix apps & tools.

Now, where is that bug?

Pedagogical implications - questions for teachers

Learner-oriented questions:

1. Do your students have **access** to a smartphone (if not, what alternatives?)
2. Do your students know their **app store**? Are they logged in?
3. Have they got their **headphones** at hand (if in class or if necessary)?
4. Have you provided **instructions**, the right **name** of the app, **logo**?
5. Have you warned them about potential **threats**?

Content-oriented questions:

1. Have you tested the app content **yourself**?
2. **Why** are you introducing the app (=purpose)?
3. Will the app content meet your students' **needs** (=what can they learn from it)?
4. Does the app content **link up** to the course syllabus?
5. What **feedback** does the app offer (=binary or targeted)? How beneficial?
6. How will your students' **react** to the app?
7. **How many** apps are you going to introduce on a course?

References (1)

Cardoso, W. (2018). Learning L2 pronunciation with a text-to-speech synthesizer. Future-Proof CALL: Language Learning as Exploration and Encounters – Short Papers from EUROCALL 2018, 2018(2018), 16–21. https://doi.org/10.14705/rpnet.2018.26.806

Cardoso, W., Smith, G., & Garcia Fuentes, C. (2015). Evaluating text-to-speech synthesizers. December, 108–113. https://doi.org/10.14705/rpnet.2015.000318

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching pronunciation hardback with audio CDs (2): A course book and reference guide* (2nd ed.). Cambridge University Press.

Chapelle, C. A. (2001). Computer applications in second language acquisition. Cambridge University Press.

Dizon, G., & Tang, D. (2020). Intelligent personal assistants for autonomous second language learning: An investigation of Alexa. JALT CALL Journal, 16(2), 107–120. https://doi.org/10.29140/jaltcall.v16n2.273

Eksi, G. Y., & Yesilcinar, S. (2016). An Investigation of the Effectiveness of Online Text-to-Speech Tools in Improving EFL Teacher Trainees' Pronunciation. English Language Teaching, 9(2), 205. https://doi.org/10.5539/elt.v9n2p205

Fouz-González, J. (2015). Trends and directions in computer-assisted pronunciation training. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English pronunciation: Trends and directions* (pp. 314–342). Palgrave Macmillan. https://doi.org/10.1057/9781137509437

Gottardi, W., Almeida, J. F. de, & Tumolo, C. H. S. (2022). Automatic speech recognition and text-to-speech technologies for L2 pronunciation improvement: reflections on their affordances. *Texto Livre*, 15.

Handley, Z. (2013). Text-to-Speech Synthesis in Computer-Assisted Language Learning. In: CHAPELLE, C. A. (Ed.). *The Encyclopedia of Applied Linguistics*. New York: Wiley-Blackwell. p. 5846–5851

Henrichsen, L. E. (2020). An illustrated taxonomy of online CAPT resources. *RELC Journal*. https://doi.org/10.1177/0033688220954560

References (2)

Inceoglu, S., Lim, H., & Chen, W. H. (2020). Asr for EFL pronunciation practice: Segmental development and learners' beliefs. Journal of Asia TEFL, 17(3), 824–840. https://doi.org/10.18823/asiatefl.2020.17.3.5.824

Jurafsky, D., J. H. Martin (2021). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (3rd ed., Unpublished Manuscript). https://web.stanford.edu/~jurafsky/slp3/ed3book_sep212021.pdf

Kim, I. S. (2006). Automatic speech recognition: Reliability and pedagogical implications for teaching pronunciation. Educational Technology and Society, 9(1), 322–334.

Levis, J., & Suvorov, R. (2013). Automatic Speech Recognition. In C. A. Chapelle (Ed.), The encyclopedia of applied linguistics (pp. 316–323). Wiley-Blackwell.

Levis, J. (2007). Computer technology in teaching and researching pronunciation. Annual Review of Applied Linguistics, 27(May), 184–202. https://doi.org/10.1017/S0267190508070098

Liakin, D., Cardoso, W., & Liakina, N. (2017). Mobilizing Instruction in a Second-Language Context: Learners' Perceptions of Two Speech Technologies. *Languages*, *2*(3), 11. https://doi.org/10.3390/languages2030011

Lieshout, C. van, & Cardoso, W. (2022). *Google Translate as a Tool for Self-Directed Language Learning*. *26*(1), 1–19.

McCrocklin, S. (2019). Dictation programs for second language pronunciation learning : Perceptions of the transcript, strategy use and improvement. 7(2), 137–157. https://doi.org/10.30438/ksj.2019.7.2.2

McCrocklin, S. M. (2016). Pronunciation learner autonomy: The potential of Automatic Speech Recognition. *System*, 57(April 2016), 25–42. https://doi.org/10.1016/j.system.2015.12.013

McCrocklin, S. M. (2014). Dictation programs for pronunciation learner empowerment. Proceedings of the 5th Pronunciation in Second Language Learning and Teaching Conference., September, 30–39.

References (3)

Moon, D. (2012). Web-based text-to-speech technologies in foreign language learning: Opportunities and challenges. Communications in Computer and Information Science, 352 CCIS, 120–125. https://doi.org/10.1007/978-3-642-35603-2_19

Mroz, A. (2018). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. Foreign Language Annals, 51(3), 617–637. https://doi.org/10.1111/flan.12348

Moussalli, S, W. Cardoso. (2019). Intelligent personal assistants: Can they understand and be understood by accented L2 learners? Computer Assisted Language Learning, 33(8), pp. 865-890. Informa UK Limited. https://doi.org/10.1080/09588221.2019.1595664

Munro, M. J., & Derwing, T. M. (2015). Intelligibility in research and practice: Teaching priorities. In M. Reed & J. M. Levis (Eds.), The Handbook of English Pronunciation (pp. 375–396). Wiley Online Library. https://doi.org/10.1002/9781118346952.ch21

Neri, A., Cucchiarini, C., & Strik, H. (2006). ASR-based corrective feedback on pronunciation: Does it really work? INTERSPEECH 2006 and 9th International Conference on Spoken Language Processing, INTERSPEECH 2006 - ICSLP, 4(May 2014), 1982–1985. https://doi.org/10.21437/interspeech.2006-543

Neri, A., Cucchiarini, C., & Strik, H. (2008). The effectiveness of computer-based speech corrective feedback for improving segmental quality in L2 Dutch. ReCALL, 20(2), 225–243. https://doi.org/10.1017/S0958344008000724

Rogerson-Revell, P. M. (2021). Computer-Assisted Pronunciation Training (CAPT): Current Issues and Future Directions. *RELC Journal*, 0033688220977406. https://doi.org/10.1177/0033688220977406

Walesiak, B. (2021). Mobile apps for pronunciation training. Exploring learner engagement and retention. In A. Kirkova-Naskova, A. Henderson & J. Fouz-González. (Eds.), English Pronunciation Instruction: Research-based insights (pp. 357-384). John Benjamins. https://doi.org/10.1075/aals.19.15wal

Yu, D., L. Deng. *Automatic Speech Recognition A Deep Learning Approach*. London: Springer, 2015.

Thank you!

Beata Walesiak
beata.walesiak@uw.edu.pl

William Gottardi
william.gottardi@posgrad.ufsc.br