Speech Recognition Basics

What is Speech Recognition?

Speech Recognition technology compares and scores speech input from students with native speaker models. If the input is close enough to a target model sentence, then recognition occurs. If the confidence level is 0.2 or above, the sentence will be recognized. DynEd has set this minimum level so that learners are not frustrated by the system being too strict. This also allows for accent variations. The great strength of SR is that it encourages students to practice speaking and to improve both their fluency and their pronunciation through increased practice.

What kind of speech input is the recognizer set for?

DynEd has set the models for English language learners who need to speak clearly and who don't speak as fast as native speakers. Native speakers who speak very fast and who don't speak clearly may not be recognized. They are not the target user.

Why is it possible to fool the recognizer?

The recognizer allows for variations in speech input. It has also been set at a low enough threshold so that learners are not frustrated when they try to follow the models. If the recognizer is set to look for random or intentional variations, then the system becomes too strict for English language learners. Please note that the SR levels vary by course and lesson. Course designers may make the system much stricter for higher level students.
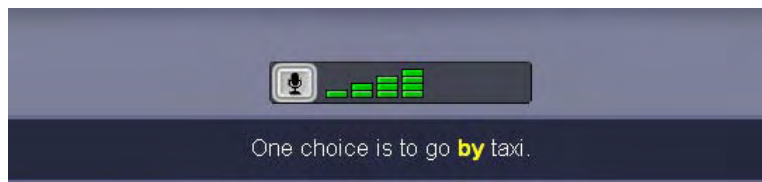
What does the speech gauge indicate?

The speech gauge indicates how confident the recognizer is that the input matches the target model. If it is not confident, it shows a single red bar and indicates which part of the sentence has scored the lowest. This helps students *focus* on improving that part of the sentence. When the sentence confidence score is above 0.3 and below 0.4, the gauge shows yellow. (See figure below).

In this example, the gauge shows yellow and indicates that the student should say "for work" more clearly. Note that this 'focus' information does *not* show when the gauge shows green bars.

What is the confidence score?

The speech gauge indicates how confident the recognizer is that the input matches the target. If it shows 4 greens, as above, then the confidence score is 0.5 or above. If it shows 3 greens, then the confidence score is between 0.4 and 0.5. The displays and confidence score ranges are as follows:

4 Greens	0.5 – 1.0
3 Greens	0.4 – 0.5
2 Yellows with focus hint	0.3 – 0.4
1 Red with focus hint	0.2 – 0.3
??? or no recognition	0.0 – 0.2

The confidence score is the score for the *whole* sentence. Please note that it is possible to have a passing sentence score even if the score for a single word is zero. This can happen if someone with good pronunciation substitutes an incorrect word. Though one word may score 0, clear pronunciation of the other words gives an overall passing score. Even native speakers may misspeak a word and listeners will know what was intended if everything else is understood.

How can a student try again?

If a student wants to record the same sentence again and have it recognized again, click on the *microphone* button *in the speech gauge.* This is excellent practice and should be encouraged. When the gauge shows red or yellow, the focus information will help the student know where to speak more clearly. The student should also listen to their recording and compare it to the native speaker model. Don't just repeat it the same way every time.

What does the data show?

Data collected over several years and in several countries shows the following:

1. Students enjoy using Speech Recognition exercises and tend to favor those lessons.

2. Students make significant gains in pronunciation and oral fluency by using Speech Recognition.

3. With frequent practice over a 3-month period, groups of students have improved their *average* recognition scores from less than 50% to more than 65% even while the language tasks became more difficult.

Common Questions About Speech Recognition

1. Why is the recognizer very strict about some words and not others?

The recognizer focuses on key words or constructions depending on the lesson type. For example, in some lessons the recognizer will focus on the difference between 'live' and 'lives' or 'do' and 'don't'. In other lessons, the focus will be on word order and not the individual words. If a key word is wrong or not recognized, even with good pronunciation of the other words in the sentence, the recognizer will not recognize the sentence.

2. Why does the recognizer show a '?' at the end of a sentence even when the correct word was spoken?

Some speakers let their volume fade at the end of the sentence. In many cases, the recognizer wants to confirm that a sentence is completely finished. Therefore, if the last word is not heard clearly, it will not recognize the sentence even if spoken clearly.

3. Sometime the recognizer shows 3 greens even when everything is clearly spoken. Why don't 4 greens show?

Speech Recognition technology still has limits. Some phonemes are not recognized as well as others. If a student gets 3 greens, that's generally good enough. DynEd continually tunes the recognizer to improve the reliability of speech scoring. Some voices are also better recognized than other voices. Adults are generally recognized better than little children. This is because the native speaker models are generally based on adult speech. When the recognizer shows 4 or 3 greens, no focus information is given.

4. Does the recognizer score for intonation and stress patterns?

These patterns are highly variable and difficult to detect. The recognizer doesn't focus on these speech characteristics at this time. However, too much or too little stress may cause distortion or the missing of a word, which will result in poor recognition.

5. Why does the recognizer work poorly in some cases or on some computers?

The recognizer requires a good microphone and not too much background noise. The student should also keep the microphone away from the mouth, since puffs of air will cause distortion. Put the microphone away from the front of the mouth, and a bit off to the side. Another problem is the microphone settings on the computer. Sometimes a microphone boost is necessary, and the default setting on the computer may need to be changed. For problems like these, please go to http://www.dyned.com/support/ . Go to the section on Voice Record / Speech Recognition.