Fundamentals of language Acquisition

Prof. BIDISHA SOM

Dept. of HSS

IIT Guwahati

Week 03

Lecture 014

Lec 14: Speech segmentation

Welcome back. Today, we will start with speech segmentation. This is our Lecture 4 of Module 3. So, speech segmentation is something that children and infants do automatically. However, this is actually quite a task. And how do you know this is a difficult thing? Just imagine yourself listening to a foreign language utterance, a speech, or, you know, any dialogue, monologue, whatever.

So, the problem here is that, unlike the written part of language, acoustic signals do not have any important markers that tell us where to stop or where the gaps between words are. So, then in terms of space, in the written version of a language, we have space to designate the distinction between two words, but that is not how speech works. In the spoken variety of language, this is not how it occurs. And then this actually is a problem, and that is why I gave you the example of the foreign language.

You need to know where one word ends and where another word begins. And this is important because just knowing the difference between speech sounds is not the same as knowing the difference between words. And that is ultimately what the infant needs to do; knowing only the sound differences will not really work. So, now the question that we are facing is: What cues do infants use to segment the speech stream, because they ultimately do? So, what do they probably use? So, there are many theories and many proposals depending on the phonological cues and other cues. So, we will look at them one by one.

One of the most common understandings is that children, including infants, use prosodic cues. Now, what is a prosodic cue? Prosodic cues refer to the rhythm, stress patterns, and intonation of speech. So, every language has its own kind of rhythm pattern, stress pattern, and also the intonation pattern. So, based on that, children probably find out the word boundary, which is one of the ideas. Because each language has its own pattern, you can figure out where the word is based on that.

So, it is the word final position or word initial position where the rhythm and stress occur that could be a very important indicator. For example, there are two main types of stress patterns: one is called the trochaic system, and the other is called the iambic pattern. So, in the case of a trochaic pattern, in the case of the English language, for example, if you follow this, it will sound like "doctor," "headache." So, the first part of the word gets the stress, the most stress. So, that is why it is called the strong-weak pattern.

The strong-weak pattern is called the trochaic pattern. So, strong followed by weak stress, again strong followed by weak stress, like that. Iambic, on the other hand, is the opposite. So, weak and strong. So, weak stress is followed by strong stress.

There are some words in English that follow this, for example, giraffe and then trapeze, and so on. So, this is how the prosodic cues work. Now, using this, infants can potentially learn how to segment the speech stream. So, if English infants assume a strong-weak combination as a word, which is a more common pattern in English, then they can accurately predict the segmentation. So, they know when the strong is followed by weak means that is one word, and then there will be a gap; then again the same pattern will repeat.

If that is how they understand, then they will be able to segment. This is known as the Metrical Segmentation Strategy. It was proposed in 1988. Now there are lots of experiments that have been carried out in this line. One of the most famous is the 15 sets of experiments done in 1999.

These studies showed that English-learning infants are sensitive to the strong-weak stress pattern and can use it to segment words from a continuous speech stream. There were a series of experiments, 15 to be precise; one of them involved 7-month-old infants, who were familiarized with words containing trochaic stress, similar to the words we saw previously, like "doctor." So, trochaic stress is characterized by a strong stress followed by a weak stress. So, the strong-weak pattern is the trochaic pattern. Children who were 7-

month-old infants were familiarized with a set of trochaic patterned words, not in isolation but within a text.

So, embedded in a text, they were listening to texts that had words in the trochaic pattern. In the testing phase, they listened to words in isolation. Some of those same words, and some were unfamiliar, and they were listening to the words in isolation. What the results showed is that they listened to familiar words longer than the new words. So, this is the comparison between the familiar words that had a tracheic pattern and the unfamiliar words that had whatever pattern, but the difference was this.

And this was taken to be proof of the children identifying the patterns. The same group also carried out other kinds of tests, one of which was familiarization with monosyllabic words and then testing them on disyllabic words. Now, monosyllables and disyllables will have different kinds of stress patterns even if the same word appears in the disyllabic form. So, king, ham, and then you are testing them on Kingdom and Hamlet. So, when it is "king" and "ham," the stress pattern will be one kind, and in the case of "kingdom," that will be different.

In this case, the infants did not show any preference for words containing the familiar monosyllabic words. The previous example had familiar versus non-familiar, the familiar one had a particular kind of stress pattern, and they showed a preference for the familiar words. Here, what they tried to check is whether it is familiarity or something else that is making them look at the familiar words for a longer time. And here the infants did not show any preference for words containing familiar monosyllabic words. The reason and the justification for this finding that was given is that it is not the sound of the word; meaning, the familiarity is not what is happening here, but the stress pattern that the child identified.

Because the stress pattern of the same thing changed in the dissyllabic format, they are now not able to identify the format and hence could not recognize the same words in the disyllabic form. However, there are problems with this kind of finding; first and foremost, this will not work in those languages where stress patterns do not provide a reliable cue for word boundaries. There are many languages in the world that do not follow a very simplistic, cut-and-dried kind of stress pattern for word boundaries. So, for example, in one of the Eskimo languages, the words may vary in stress pattern under different conditions, depending on different positions and other cases. So, the same word may have different stress   patterns.

So, then it is not a very reliable cue; stress, in this case, is not a very reliable cue for identifying the word boundary. So, there is similarly a large number of languages that do not have word-initial or word-final stress patterns that can be used reliably for word segmentation. Because languages differ across the world, each language has its own pattern. In some cases, like the English stress pattern, it is a very reliable cue; in other languages, it may not be a very reliable cue. So, for those languages, Cutler suggests infants might be using other strategies.

What other strategies are we discussing? There are many others. So, another idea that has been put forward is called phonotactic regularities. Now, what previous one was prosodic pattern, now this is phonotactic regularity. Now, phonotactic regularity refers to the system that determines which sounds and sound combinations are allowed. So, the system of sound and sound combinations that not every language has, you know, the same kind of sound combinations allowed

.

For example, in English, the combination of V, Z, and G is not allowed, but this is possible in Russian. Or if you know, the most famous example could be Polish. So, in the case of the Polish language, the kind of consonant combination that is allowed is not permitted in most languages of the world. English does not allow, nor do any other Indian languages allow that kind of combination. So, this is what phonotactic regularity means.

So, because of this pattern, if the child is aware of it, then they will probably be able to understand. Similarly, the /br/ combination is used more often in the word-initial position in the English language, and /nt/ more often in the word-final position. So, this kind of combination, what is allowed and what is allowed in which position, is what creates our phonotactic patterns in a language. So, if infants are sensitive to the combination and it is in an allowed position, then they can use it to segment words. So, there have been a lot of studies that have looked at this idea, and they have found proof for that as well.

One of the most well-known studies is the one carried out by Mattys et al. So, the study looked at whether infants are sensitive to phonotactic regularities. So, what they did was present them with syllable sequences that varied in probability of their occurrence within words. So, the paradigm used here was head turn preference, which we have already discussed before. So, based on this paradigm, they created some nonsense syllables, nonsense words, non-words, and they had this CVC CVC pattern of syllables and bisyllabics;   they   were   all   by   syllabics   CVC   and   CVC.

So, two syllables here, and they were in this consonant-vowel-consonant, consonant-vowel-consonant cluster. This is how the patterns were created. So, there was a list of bisyllabic words that were created, and all were non-words. And what they did was the primary, the variable number 1 was that the non-words bore primary stress in either the first or the second syllable. So, the stress patterns were changed.

So, it is either basically trochaic or iambic. So, a weak-strong or strong-weak kind of pattern. So, syllables differed in this. Bisyllabic words differ, non-words words, either on a triambic pattern or an iambic pattern. The second variable that they tested was the phonotactic nature of the cross-syllabic C C cluster.

So, you see C C V C C V C. So, that is this C C cluster that is coming between; that is what was variable 2. They were trying to check, by manipulating the kind of clusters, they wanted to see if this works to identify word boundaries. So, clusters had a high versus low probability; the slash means "versus" here, of occurring in either a within-word or between-word position. So, they also manipulated those syllables depending on the clusters; some clusters had a higher probability of appearing in a word inside a word, and others had a higher probability of occurring between two words, so inter-word location. So, for example, /ngk/ is more common within words at syllable boundaries, but /ngt/ is more common between words at word boundaries.

So, if a /ngk/ is a cluster, that means it is a single word. However, if you hear the /ngt/ cluster, that means it is a boundary-marking cluster at the end of the word. Now, this is how they manipulated the stimuli. What they found out was that the subjects were 9-month-old infants, and the infants showed different kinds of behavior patterns with respect to the two types of clusters, roughly meaning that they understood the difference. So, they showed a pattern of proof of understanding the differences, and they could cluster them depending on the cluster that could help them get the word boundary or the phonotactic regularities.

The third kind within this larger category, that we can call phonological bootstrapping, is called the allophonic variation. We all know allophones. The different modifications that a phoneme takes based on the acoustic environment. So, the position of a word, or a sound based on what precedes or follows it, is how the sounds change. We give the example of /pa/ sounding like /pha/ in the initial position of a word in English or /ta/ becoming /tha/ in the initial position of a word; that is an allophone.

So, this variation refers to the different pronunciation of the same phoneme in different acoustic environments that are in different positions in the word. So, like /pa/ and /ta/, etc. So, this is a very subtle case of correspondence between the manifestation of a sound and its position. So, in English, if you are hearing the /tha/ sound, that means it is at the beginning of a word. So, you will automatically know the word boundary.

So, again, studies have been carried out, and they found that infants can differentiate between aspirated and unaspirated sounds, something like /pa/ and /pha/. So, they set out to find out if infants can use this to locate the word boundaries. So, first they found out that the infants are sensitive to aspiration as a marker of difference between sounds, and then they wanted to see if this understanding leads to understanding word boundaries. So, again by the same group, they looked at the used condition head-turn procedure as the paradigm. So, speech sounds play only when the infant turns their head towards a flashing red light above the speaker; you know the drill.

So, the same thing. So, in the familiarization phase, infants heard the word "nitrates" or "night rates." So, what is happening here is that the same words put together are almost the same words put together in a single word versus two words, but the children have never heard both of them together. So, they were hearing either nitrates or night rates. Along with many other filler words like "doctor," "Hamlet," etc. In the test phase, they heard passages that included both words.

So, initially they heard only the words; that was the familiarization process, and then in the test phase, which is the second phase, they heard those words embedded in a sentence context. So, sentences like "businesses try to use night rates to send their packages". So, for example, we have these embedded night rates here. Similarly, this starts here at the beginning. Similarly, again at the end, you can see that they have also modified the position of the words they were trying to check.

So, businesses try to use nitrates to send their packages; nitrates are something that everybody needs, and so on. What did they find? They found that infants are not expected to understand; of course, infants are not expected to understand the sentences. However, what they discovered was that the children who had been familiarized with nitrates listened longer to the same word in the phrase, and the same was true for night rates. The effect was seen in the case of 10 and a half-month-old babies, but not in the case of 9-month-old

babies. So, there was a slight difference between the age group of the infants that took part in this experiment, and based on the age, there was a slight difference in the results that they had.

So, 10 and a half-month-old kids could find the difference, but then the 9-month-old babies could not find the difference, which means that this understanding, this sensitivity, grows with time. So, that is what they concluded: the sensitivity to allophonic cues develops between 9 and 10 and a half months of age. Because, in this case, you have seen that they have changed the position of the words based on what is embedded in the sentences. So, these allophonic cues are fine; they understand, but it takes a little bit of time to develop. So, we see that infants are sensitive to prosodic cues and allophonic cues.

But being sensitive and using them to segment words are not the same thing; that is what our primary issue is here. Also, overreliance on these cues can lead to mis-segmentation. If the infant completely bases his or her understanding of segmentation on these kinds of cues, then this might also lead to problems that can be called overgeneralization because this will not always work. For example, aspirated /ta/ occurs not only in the word-initial position but also internally, as in atomic. Now, atomic is not here; the sound here is not in the beginning.

However, even here, the /t/ sound will be aspirated. So, these are certain cases where the set rule does not work. So, these are exceptions, but if we use a generalized pattern to understand this, there will be an error; that is what we are trying to say. Hence, babies also need to know where the cue is useful for segmentation and where it is not. Also, the pattern of cues differs across languages

.

So, one particular format will not apply to all languages. Now, this brings us to a very curious kind of problem. In order to know the relevant cues for a specific language, infants need to learn a certain amount of words. Only after you have some amount of data can you imagine how machine learning and all these artificial agents are trained. The more data you train it with, the better the outcome; the better the test result, the better the fit will be.

The same thing is applied to humans. So, in order for the child to understand where the cue is important, where it is not, where it works, and where it does not, they need to know

at least a few words; they need to have some understanding of the words in the language. And to know the words, they need to learn word segmentation and word boundaries. So, this is a Catch-22 situation. So, for example, in order to use the trochaic stress pattern to segment English words, English infants need to know that trochaic stress is specific to this language.

That is what can be understood as an example. So, one possible answer to this problem is that they start by using different cues that help them learn a few words. Once a few words are in the vocabulary, by different cues we probably mean a set of cues. It could be allophonic; it could be prosodic cues; you know it could be any of the three that we have already talked about. So, once those cues are used, the child gets to understand a few words, has kind of figured out a few words, and then, using it as a base, they build up the later understanding of the rest of the vocabulary. So, they use those to learn the language-specific cues like stress patterns, etc.

Then those cues are used for learning more words. So, you learn a few words, understand the pattern, and use the pattern to understand more words; that is one of the hypotheses. So, based on this idea, the theory of the isolated word was proposed. What does it mean? What does "isolated word" mean? Children actually learn isolated words in the initial phase of their lives. How to go about finding whether this works or not? There was a rather well-known study carried out in 2001. where they explored the idea of studying what English-learning children actually hear

.

So, for these studies, they analyzed transcripts of mothers talking to their children. Because the primary thing they were trying to find out is whether they actually learn isolated words. Are isolated words a part of their ambient language? How many isolated words do they hear in the early stages of their life? If they hear an adequate amount of isolated words and eventually identify an adequate number of isolated words, this could be taken as proof of a correlation between the two. That was the idea behind this experiment. So, the mothers with their first-born infants under 9 months old were chosen.

This duo was visited at home every two weeks while the infant was between nine and fifteen months of age. So, starting at 9 months until the child was 15 months of age, this experiment went on. The experimenters will visit the home where the mother and child will be talking; the mother will be talking to the child, and that will be recorded. They had a

total of 14 visits; the mother's speech with the child was recorded as she went about her normal life.

Sessions lasted for 90 to 120 minutes. Out of that entire session, only the middle 75 minutes from 3 to 4 sessions for each child were used for the study. Infants aged before 12 months were called the early transcript, and those after 14 months were called the later transcript. So, the first phase and the second phase are completed. So, before one year and slightly after one year. Early transcripts were used to assess the maternal speech to the infants, whereas the later transcripts were used only to assess the infants' productive vocabulary.

Because you see, this is after 14 months, so this is the time after 1 year when children are already using words, at the single-word or two-word stage, depending on the child. So, that is when they were looking at the output, and in the stage before that, from 9 months to 12 months, they were checking the input that they were getting in terms of the mother's speech. The early transcripts were analyzed to determine the frequency and diversity of isolated words in maternal speech. And isolated, how do they define isolated words? Isolated words are those that have a minimum of 300 milliseconds gap.

This is crucial. So, there has to be a gap of at least 300 milliseconds before and after the word. Those will be considered isolated words. That means that an adequate cue is given to the child that this is a word in itself. This is not part of another word. 300 milliseconds is a lot of time when you are talking about the normal speech rate.

So, that is how they put it too. So, they reported 9 percent of isolated words among all the utterances. This is not a very less amount. There was plenty of use of isolated words that the infants could hear at that stage of their lives. It was also found that those words were the first words learned by the infants. So, when they compared this with the output, speech production by the children and the infants, they found that there was a strong correlation between the words that they produced and the isolated words that they heard in the first stage of life.

This result showed that a child's ability to produce words at 12 months was correlated with the frequency of isolated words produced by the mother. Hence, they concluded that it is possible for infants to begin to learn words from the isolated words that they hear. And on the basis of which they can figure out the other phonological, phonetic, and phonological cues, they go about segmenting further words; that is the idea. However, there are

disagreements with this theory as well. One of the problems pointed out by the other group is the difference between isolated words and isolated phrases.

Sometimes the phrases can also be considered as words. This is quite a common finding among infant speech patterns. That the words "mummy" and "mummy hen" might be considered similar. So, both are considered words; at that stage are children able to differentiate between short phrases and words? That is the question here. So, if they do not know that difference, then they will learn a lot of phrases as words. So, that in itself is not a big problem since that is how they actually learn, but when it comes to word segmentation, that is a problem.

Because the phrasal structure will have a different kind of stress information, whereas the words will have a different kind of segmentation. Because we started with the idea that words would have prosodic cues on the basis of which you learn more words. Now, if children are taking phrases as words, then they will be following the wrong prosodic cue to segment other words. This is where the problem lies, because the stress pattern between words and phrases changes. Now, there is another possibility that has been put forward to take care of the previous problems, which is called transitional probabilities.

Now, this one takes care of the statistical understanding; it looks at the whole problem thinking in terms of statistical probabilities. What are the probabilities of certain kinds of sound combinations appearing in word-initial, word-final, and word-boundary positions? If a child has that kind of understanding, then they can use it. So, the idea here is that the infants are clever statisticians who can keep track of the probabilities of sounds co-occurring and what kind of sounds can co-occur. For example, if one hears the syllable /pri/ in English, what are the probabilities that the next syllable will be /ti/? Because in English, "pretty" is a very common word, a very high-frequency word, and it is also quite commonly used by mothers with infants, you know, "This is a pretty flower," "This is a pretty dress," kind of thing.

So, this is very common. As a result, this is something that the children are probably expected to learn. So, they know that /pri/ can be followed by /ti/. But the probability of /ti/ being followed by /beb/ is very low, as there is no word like /tibeb/ in English; that is what the idea is. So, unless it occurs only in a phrase like "pretty baby.

" So, where /ti/ and /beb/ come together and create a phrase in some sense. In the continuous speech stream, this might look like a combination. So, the transition between /ti/ and /beb/

happens only between words, not within. However, the transition of /pri/ and /ti/ happens within words. Now, this kind of statistical understanding of which probabilities are higher for word boundaries and which are not, is expected by the child according to this theory.

So, the transitional probability discusses this. So, if infants are capable of understanding this, it can be used as a cue to segment words. This kind of sounds almost improbable, because a child who is not even able to speak properly and is not able to segment words, assigning them a lot of statistical potential, might be a little far-fetched, but there seems to be some proof of it as well. So, there were some studies where one of them is that 8-month-old English infants could identify transitional probabilities between syllables in a continuous stream of nonsense speech sounds.

This is what they found. There are many other studies that have followed up on this topic. We are not going into the details here. I will add the references you can look up. So, there are adequate amount of studies today available that prove that probably, the mind can work on a statistical model as well. The infant brain can actually figure out what kind of possibilities are more probable compared to other possibilities.

These are one. The other kinds are all tied to some sort of, you know, either phonetic or phonological ideas. So, this is what we can call, and we can consider all of them in some sort of phonological bootstrapping. Because they are trying to use the phonological or phonetic properties of the language cues they receive, we think they are trying to segment the continuous speech stream into words and find word boundaries. The other kinds of theories talk about the possibility of using meaning as a source to help infants.

So, this is also called semantic bootstrapping. Semantic bootstrapping is not used only for word boundaries, but in syntax as well, as we will see later. But, roughly we can say that semantic bootstrapping is also one of those concepts; when we use meaning, we can call it semantic bootstrapping. For example, Barbara Lust talks about: what if language is not learned perceptually, but conceptually? We are talking about perceptual understanding; we are talking about the acoustic cues and other, you know, phonetic and phonotactic regularities, and so on. So, those are perceptual properties, but what if children learn language conceptually? So, when you talk about concepts we actually think about meaning.

So, that is what semantic bootstrapping is all about. So, this essentially says that infants probably determine the meaning of the language first, and with this, they finally crack the rest of the word. So, how do they figure out the meaning? It is actually not very difficult for infants to understand the meaning of at least a few words. Which words? The words that are typically used around them very often in their childhood, and the designated object

is also present in the same environment. This happens with some words, like "mummy," for example. This is what the children learn very early in life because the object, the referent, is always present.

So, the mummy is always there. So, these are some other words like this that can be understood easily and learned easily because the object of reference is already present there. So, this could be the beginning of using concepts to understand and crack the rest of the language; that is what the idea is. Based on this, the multiple cues approach has been proposed. So, this idea says that children actually use a lot of different kinds of cues, not just one single cue, and one of them could be meaning. So, a new approach by Lieven and Ambridge suggests that infants need multiple cues to help them sort out the issue at hand.

When infants hear a sound pattern with an object they know well, they tend to learn it as a word rather than waiting for the, you know, phonological cues and so on; they match these two very early in life. So, you know certain things like mummy and others, so that sound and object pair is always present. And hence the chances of learning those things as one entity, as one particular thing, you know, word or whatever they understand as one unit, is quite high. So, this way, meaning help can help learn the initial words, and then stress pattern-related cues can help them learn more words. So, that is the approach that the multiple cues approach talks about, which states that both semantic and phonetic and phonological cues will come together to help children understand or figure out the word boundaries.

This is where we complete lecture 4, and in the next lecture, we will talk about speech production by infants. Thank you.