Development of Speech Perception

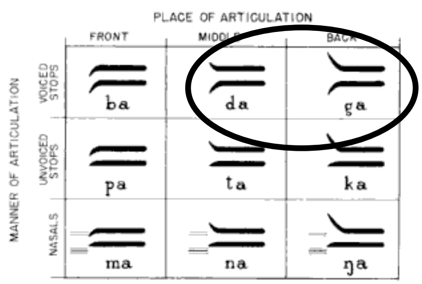
Issues in the development of speech perception

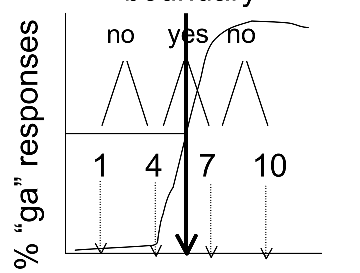
- Are the mechanisms peculiar to speech perception evident in young infants?
- Do infants, children and adults use the same information to identify speech?
- How long does it take for speech perception to mature?

Overview of lecture

- Development of speech discrimination
  - Infants
  - Preschoolers
  - School-age children
- Effects of experience on speech discrimination

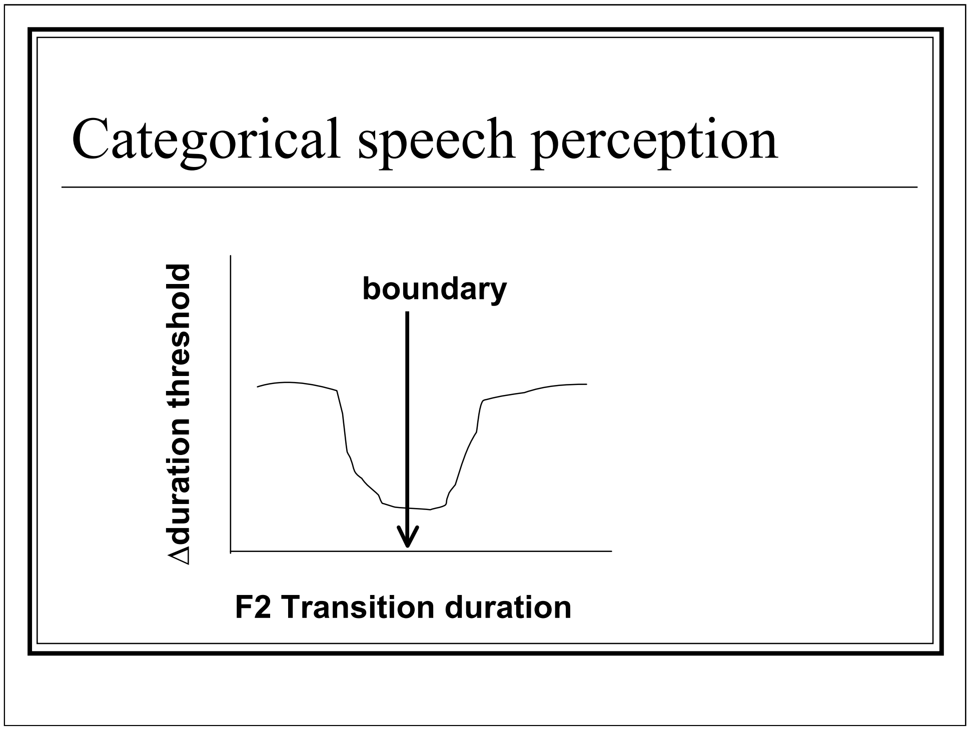
Categorical speech perception



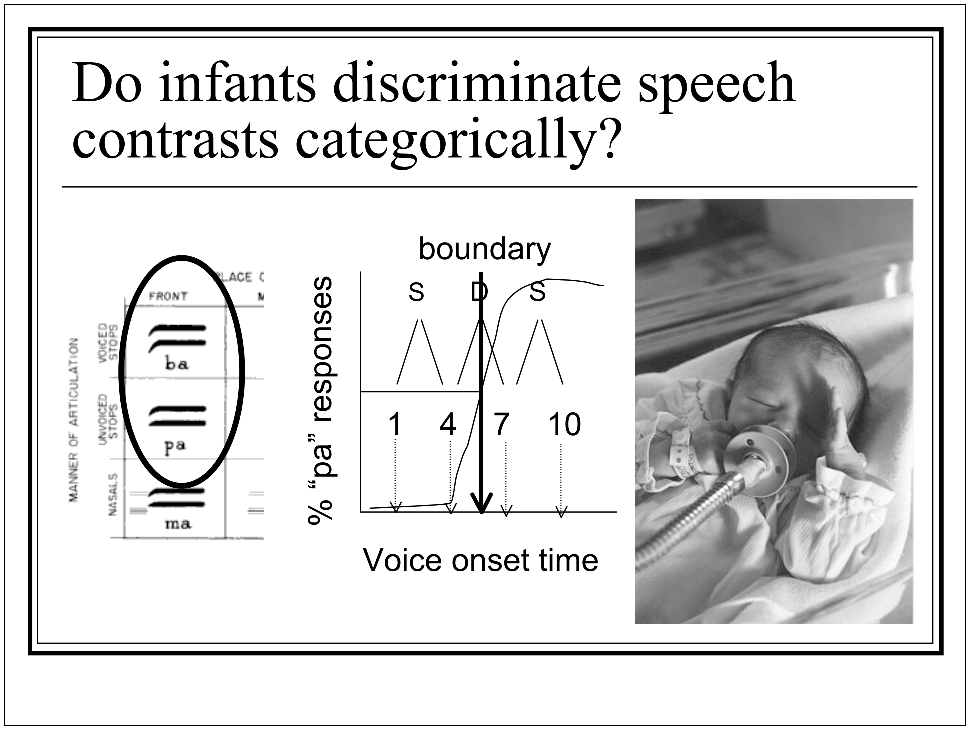


▪da-ga discrimination demonstration

One of the “mechanisms peculiar to speech perception” was thought to be categorical perception. The basic idea in categorical perception is that listeners can only discriminate between sounds that they would give different labels to. So take for example, the distinction between /da/ and /ga/, two voiced stops distinguished by place of articulation. Acoustically, these two consonants are distinguished by the duration of the formant transition to the following vowel, with /da/ having a shorter transition than /ga/. So if I construct syllables with gradually increasing formant transition duration, I would find that as the the duration increases, people would be more and more likely to label the sound as /ga/, as shown in the graph. I can define a sort of threshold--say that transition duration at which listeners say /da/ half the time and /ga/ half the time-- this is referred to as the boundary between the two phonemes. Now if I take 4 different syllables, early spaced in transition duration, with two on one side of the boundary and two on the opposite side of the boundary, I will find that people can discriminate between the two syllables that fall on different sides of the boundary, but not between syllables that fall on the same side of the boundary. People give the syllables on the left side of the boundary the label /da/? And syllables on the right side of the boundary /ga/. They can tell the difference between the syllables they assign different labels to, but not between the syllables they assign the same label to, even though the syllables on the same side of the boundary are acoustically as different as the syllables that fall on opposite sides of the boundary.

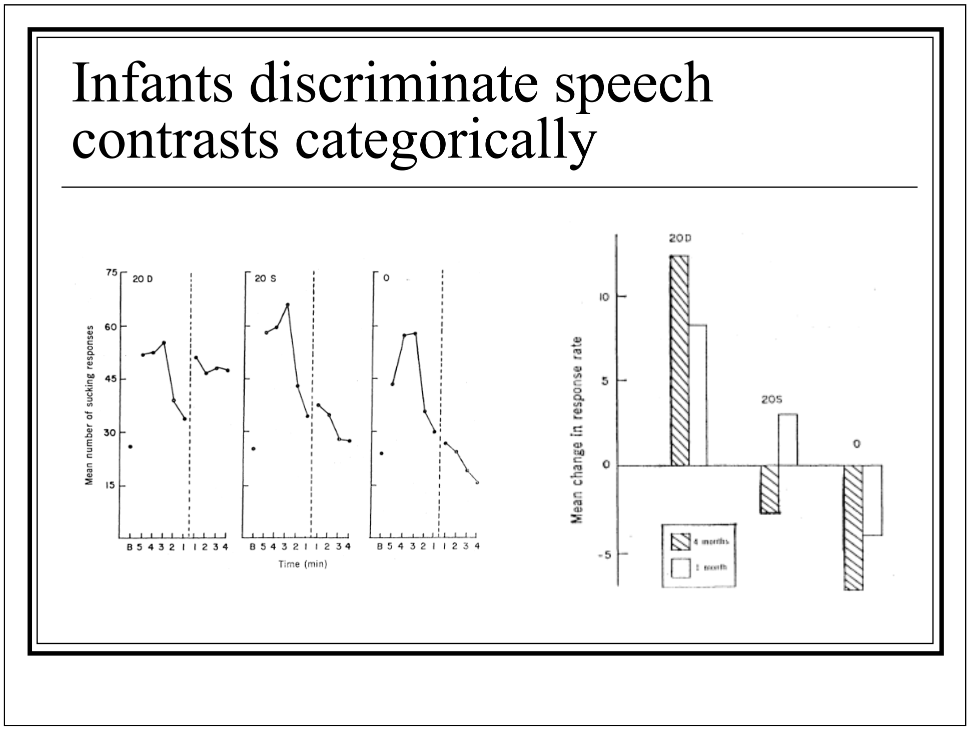


Another manifestation of categorical perception is that if I measure discrimination thresholds for the acoustic dimension under question (e.g., what is the difference between two syllables in transition duration that a listener can discriminate 75% of the time), thresholds will be lower in the vicinity of the boundary than within the speech categories.



One of the most important studies of infant speech perception was carried out by Eimas et al. in 1971. They looked at infants’ ability to discriminate the syllables /pa and /ba/, which differ in the voicing feature. One could construct a series of syllables with voice onset time--the interval between the release of the stop and the onset of voicing. For adults, when the voice onset time is greater than about 25 msec, they label the syllable /pa/ and if it is shorter than 25 msec, they label the syllable /ba/. So Eimas et el used the high amplitude sucking procedure to test discrimination of syllables that differed in voice onset time. They chose two syllables on the /ba/ side of the boundary, and two syllables on the /pa/ side of the boundary. The syllables labeled 4 and 7 in the graph, fall on opposite sides of the boundary and are given different labels by adults (D is for different) The difference between the syllables 1 and 4 in voice onset time is the same as the difference between 4 and 7, but both 1 and 4 are labeled /ba/. Similarly the syllables 7 and 10 differ from ach other in voice onset time by the same amount as syllables 4 and 7, but they are both labeled /pa/ by adults. (The S is for same). So in the Eimas et al experiment, some babies heard syllables 1 and 4, some heard 4 and 7 and some heard 7 and 10. If the babies are like adults, we expect that they will discriminate 4 and 7, but not 1 an d4 or 7 and 10. The infants were 1 and 4 months old.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*, 303-306.



Eimas et al.'s results are shown here. "20" is the difference in VOT between the two syllables, and D is for the syllables on opposite sides of the adult boundary, while S is for the two pairs of syllables that fell on the same side of the boundary. 0 is the control condition-- no syllable change was presented. Each graph on the left shows sucking rate over trials. For the trials before the dashed line, the infant is hearing the same syllable over and over again when he sucks on the pacifier, and for all of the subjects, the sucking rate goes down, or habituates, over trials. The syllable is changed at the point in the experiment where the dashed lines is inserted. For the subjects in the D condition, sucking rate increased, indicating that they heard the change in the syllable. For the subjects in the S condition, there was no increase in the sucking rate when the syllable changed. , and for the subject in the control condition, there was no increase in the sucking rate at that point in the experiment. The graph on the right shows the results separately for the 1 and 4 month olds, here just the change in sucking rate from the last habituation trial to the first change trial. The 4-monht-olds are shown in striped bars and the 1 month olds in unfilled bars. Notice that both age groups showed increases in sucking rate following the syllable change in the D condition, that that there was no significant change in the S condition. Sucking rate just kept going down in the control condition.

Eimas et al. argued that this result meant that even young infants perceive consonants categorically, and that speech must have some innate special mechanisms built into the brain that allow young infants to do this. Although subsequent research has demonstrated that

Infants discriminate many phonetic contrasts

- Consonants
  - categorical
  - voicing, place, manner
  - Initial, medial, final position
- Vowels
  - continuous (not categorical)
  - Despite variation in intonation or speaker
- Non-native phonetic contrasts

Many subsequent experiment s confirmed Eimas et al.'s findings and showed that they generalized to consonants that differed in voicing, manner and place of articulation, and to consonants at all positions within a syllable. Like adults. Infants' perception of vowels was not categorical , and even 2-month-old infants' can ignore differences in intonation and speaker, while responding to phonetic changes. An interesting finding to which we will return in the next lecture is that infants who had never heard a language could discriminate sounds that occurred in that language, but not their native language, even though their parents could not.

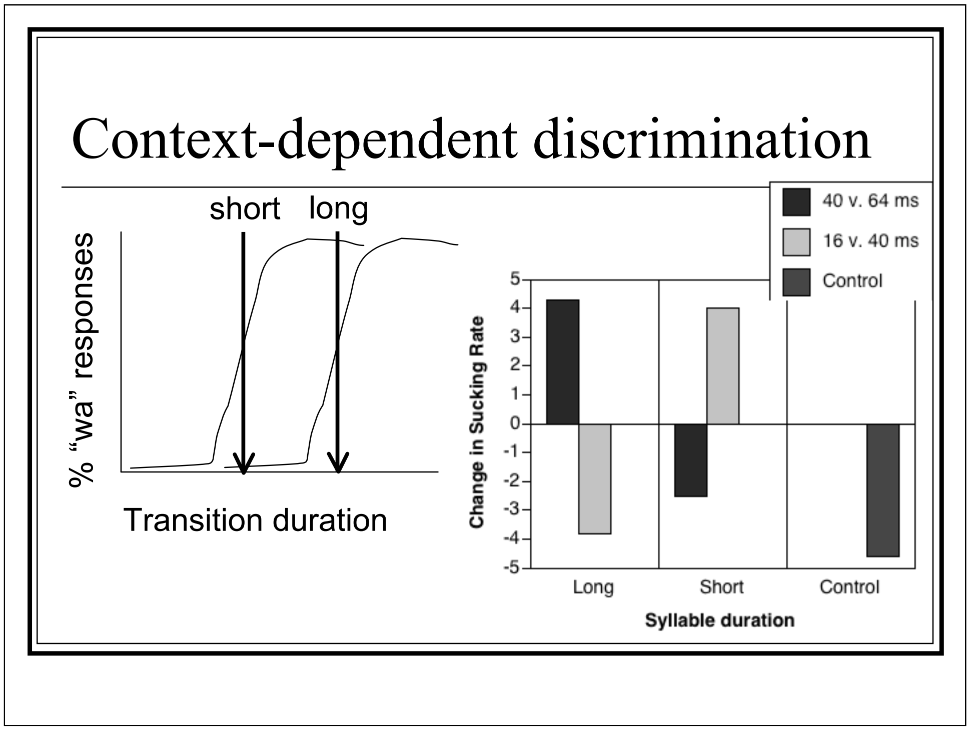
Bertoncini, J. (1993). Infants' perception of speech units: Primary representation capacities. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 249-257). Dordrecht: Kluwer.

Eilers, R. E., Bull, D. H., Oller, D. K., & Lewis, D. C. (1984). The discrimination of vowel duration by infants. *J. Acoust. Soc. Am., 75*, 1213-1218.

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Percept Psychophys, 16*(3), 513-521.

Eimas, P. D. (1974). Linguistic processing of speech by young infants. In R. L. Schiefelbusch & L. L. Lloyd (Eds.), *Language Perspectives-Acquisition , Retardation and Intervention* (pp. 55-73). Baltimore: Baltimore University Park Press.

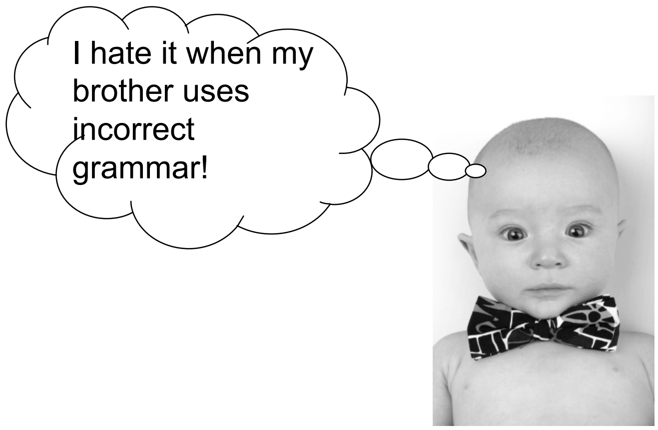
Eimas, P. D. (1975). Auditory and phonetic coding of the cues for



Eimas and Miller showed that 4-month-olds' phonetic boundaries shifted with the phonetic context as they do in adults. If a listener is asked to distinguish /ba/ from /wa/, an important acoustic cue is the duration of the formant transitions. If I present fairly long syllables (200 or 300 msec), I will find that as I increase the duration of the transition, that the listeners are more and more likely to identify the sound as /wa/. However, if I present short syllables, as would occur in s a fast speaking rate, the boundary at which people shift switch from labeling the syllable as /ba/ to labeling it as /wa/ shifts to a shorter values. This is a context effect. Eimas and miller tested 4-month-olds using ht high amplitude sucking procedure for short syllables and long syllables. They used two different pairs of syllables in each condition. One pair had transition durations of 40 and 64 msec. For adults listening to slow speech (long syllables) the boundary between /ba/ and /wa/ is between 40 and 64 msec. The other pair of syllables had transition durations of 16 and 40 msec. For slow speech, both of these would be called /ba/, but for fast speech, the boundary between /ba/ and /wa/ falls between 16 and 40 ms. So the question is which pairs do infants' discriminate and under which conditions. The results are shown on the right. When the syllables were long, infants discriminated 40 from 64 msec, but not 16 from 40 ms, but for short syllables, they did not discriminate 40 from 64 msec (both /wa/) fro adults, but they do discriminate 16 from 40 ms. So infants apparently make the same sorts of adjustments in their perception to accommodate changes in speaking rate as adults do.

Eimas, P. D., & Miller, J. L. (1980). Contextual effects in infant speech perception. *Science, 209*, 1140-1141.

Does anything about speech perception develop after birth?



These are just a few examples of the sorts of findings that bolstered the idea that speech perception is quite mature early in life. The question might be whether anything develops in speech perception postnatally at all.

One question is how it could be that young infants have immature auditory processing (frequency representation and perhaps intensity representation and temporal representation). These are three points that bear on that question. First, as far as we can tell, the most basic aspects of hearing are mature by the time that an infant is 6 months old. Many of the studies tested infants older than that, so there is nor conflict between what they are finding and what the hearing studies show. I will talk about some speech studies of infants younger than 6 months that suggest that the hearing immaturity may influence speech perception in that age group.

The second point is that there are many cues that adults can use to discriminate a phonetic contrast. Infants may be making the discrimination based on cues that adults don’t’ typically use, if they have trouble using the adult cues because of immature hearing.

Finally,when people have tested infants, they haven’t used small differences between sounds or subtle distinctions. They chose fairly large differences and sound that are easily discriminated by adults. Such stimuli don’t really tell us about the limits of what infants can do.

A test of the detail in acoustic representation of speech

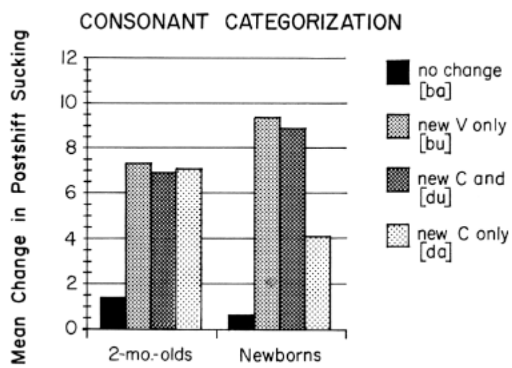
Test condition	Preshift set	Postshift set	Change
bu	bi, bɚ, bo, ba	bi, bɚ, bo, ba, bu	new vowel only
du	bi, bɚ, bo, ba	bi, bɚ, bo, ba, du	new consonant and V
da	bi, bɚ, bo, ba	bi, bɚ, bo, da	new consonant only
control	bi, bɚ, bo, ba	bi, bɚ, bo, ba	none

Bertoncini et al. (1988) investigated whether infants who listened to a set of CV syllables noticed when a new syllable was added to the set. They used the high amplitude sucking paradigm. They manipulated “how different” the added syllable was. It could be just like the old syllables except that it contained a vowel that wasn’t in the old syllables, or that it contained a consonant that wasn’t in the old syllables. In a final condition, the new syllable contained both a new consonant and a new vowel. The syllables are listed in this table. In all the conditions the infants heard “bee”, “baa” (as in bat), “both” and “bah” repeated in random order when they sucked on the pacifier. Then once the infant had habituated to those sounds, the syllable “boo”, the syllable “doo” or the syllable “daa” (as in bat) was added to the series. A control group of infants did not hear a new syllable. Remember that vowels are relatively intense, long duration events compared to consonants. So it might be that infants would hear the vowel change but not the consonant change.

Bertoncini, J. (1993). Infants' perception of speech units: Primary representation capacities. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 249-257). Dordrecht: Kluwer.

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *J Exp Psych [Gen], 117*(1), 21-33.

The acoustic representation of speech becomes more detailed

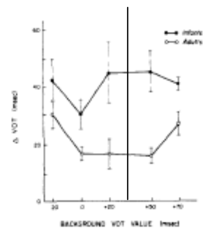


The results are shown in this figure, for 2-month-old infants and for newborns. Each graph shoes the change in sucking rate after the change in the syllable set. The black bars are for the control group. Light gray bars are for the new vowel, dark gray are for the new consonant and the dotted bars are for both vowel and consonant change. Look at the newborns first. The control group does not increase its sucking rate and the group that heard only a consonant change did not increases their sucking rate significantly, but the groups that heard a vowel change--whether or not it was combined with a consonant change--increased their sucking rate. This suggests that the newborns heard the “big” change (vowel), but not the more subtle change (consonant). By 2 months, though, infants responded to all the changes, suggesting that they could hear the consonant change as well as the vowel change.

The idea is that the newborn doesn’t hear the fine details in speech, but that by two months those details are heard. Although similar studies haven’t been done with even more subtle changes with older infants, one would predict that at least for the first 6 months of life, one would see continued improvement in the detailed representation of speech as infants’ basic auditory capacities mature.

Infants' VOT threshold

- /ba/ v. /pa/
- 5.5-11.5 months and adults
- Conditioned head-turn procedure
- Adaptive thresholds



Remember that one of the characteristics of categorical perception is that the threshold for discriminating differences between sounds is better near the category boundary. Aslin et al. measured the VOT discrimination thresholds in infants and adults to see if that characteristic of categorical perception was evident in infants. They tested discrimination along the /ba/-/pa/ continuum using a conditioned head turn procedure and an adaptive psychophysical method (VOT was increased and decreased according to the subject's responses). The subjects were 5.5 to 11.5 months old or adults.

The results are shown in the graph on the right-- ΔVOT threshold as a function of VOT (I.e., threshold for discriminating changes in a syllable with a a VOT of 25 msec or in a syllable with a VOT of 50 msec, etc.) The results for the adults are in the lower curve. VOT threshold is higher at the ends of the continuum than in the middle--as you would predict for categorical perception. The infants' thresholds are not as good as adults', but interestingly, they do best around a VOT of 0 msec, and they are least adultlike at discriminating changes at 20 and 50 msec, where adults are quite sensitive. So because infants are more like adults at some VOTs than at others, we might argue that they can't be doing especially badly in some conditions because of something like attention or memory, because they are able to perform the task better in other conditions. In other words, these results suggest some development in the ability to use VOT to discriminate speech sounds.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Dev., 52(4)*, 1135

How do infants figure out where one word ends and another begins?

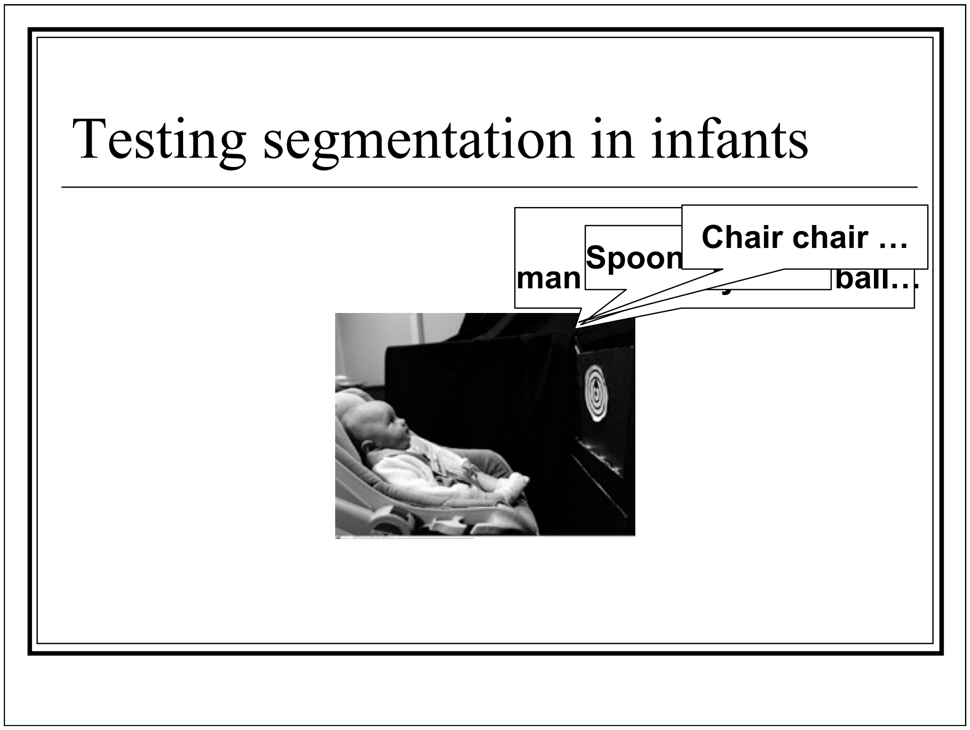
Evenininfantdirectedspeechfewutterance
sconsistofisolatedwordsinanalyzingtheinp
utheardbyaninfantbetweensixandninemo
nthsvanderwiierfoundthatexcludinggreeti
ngsvocativesandfillersonlyaboutsevenper
centofthespeechconsistedofonewordutter
ancesthustolearnthewordsoftheirnativela
nguageinfantsmusthavesomeabilitytoseg
mentwordsfromfluentspeechasdiscussed
earlierpausesrarelyoccurbetweenwordsin
fluentspeech.

Recent research has focused on an especially important and interesting area in the development of speech perception-- the development of the ability to segment running speech. When infants hear speech, they don't hear people saying "bah bah bah pah"; they hear people saying "Is that Mickey Mouse sitting next to you on the bed?" The problem here is that if we look at the acoustics of such running speech, there are no pauses between words to tell us where one word ends and another begins. You may know how hard it is for someone learning a second language to figure out this problem. So when do infants start to be able to do this?

Cues that adults can use

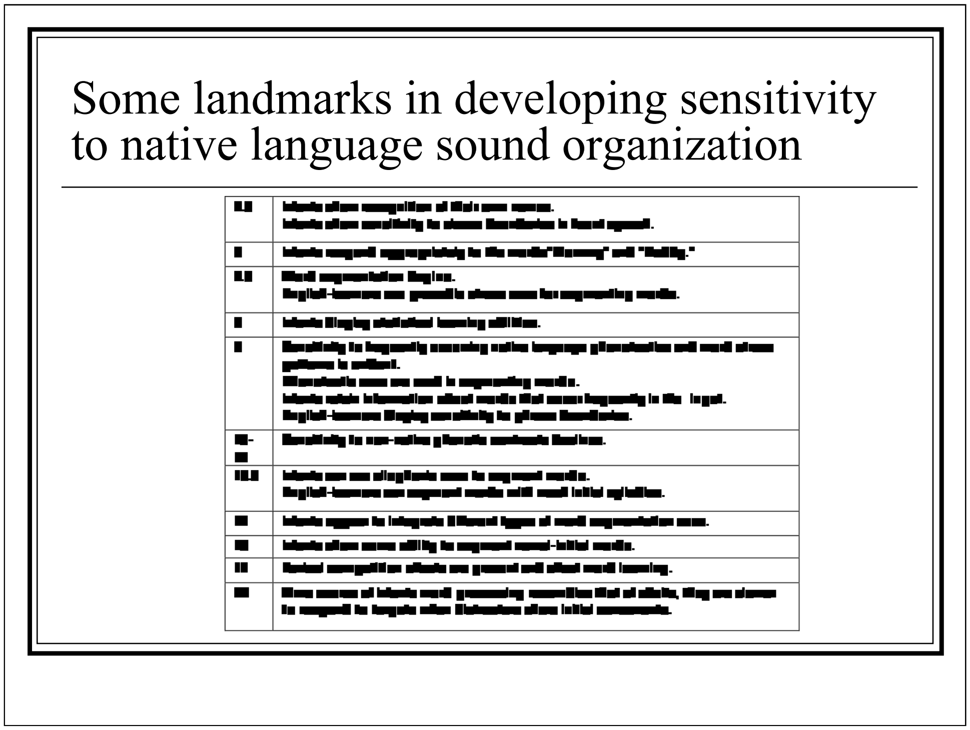
- Phonotactic regularity: “slo”, not “slx”
- Allophonic variations: [t] starts, [t^h] ends
- “correlations” between sounds: b+a+L happen together a lot.
- In some languages, words tend to follow a certain stress pattern (e.g., strong-weak in English).
- Once you know some words, you can separate those from other words.

There are several sorts of information that adults can use to segment running speech into words. For example, in any language, there are phonetic sequences that are “allowed” and others that rarely occur within a word. For example, words might start with the letters “smo”, but not with the letters “smr” in English. So if you hear “chasmreader”, the break between words is more likely to be between the m and the r. Another cue to segmentation is that the same phoneme may be pronounced differently when it occurs in different locations within a word; for example /t/ tends to be aspirated when it occurs at the end of a word in English. Listeners might also notice that the same sequence of sounds tends to occur repeatedly in running speech. Sounds that always occur together in the same order might be words. And finally, once you know a few words, then you know when those words begin or end, so you can separate the words before them and after them.



Peter Jusczyk and his colleagues carried out a long series of studies investigating infants’ abilities to use these cues. These studies generally used visual fixation/habituation methods. Basically, the idea is to present infants with ongoing speech with certain words embedded in it, and then to test whether infants recognize those words when they are presented in the second stage of the experiment.

A variation on this procedure is to manipulate the stimulus in some way and then to see whether the infants look longer for the normal than for the manipulated stimulus.



The results of all these studies are summarized in this table from a 2002 review chapter. Even young infants seem to be sensitive to some types of information in speech. For example, 4.5 month olds are able to pick out their names from ongoing speech, but it isn't until 6 months that they can pick out Mommy and Daddy. True word segmentation begins around 7.5 months. They also notice if a pause in a sentence comes in the middle of a clause instead of at the end of a clause; they look longer if pauses are inserted at clause boundaries in running speech than if the pauses are randomly inserted at other points.

By 8 months, infants show evidence of learning that certain combinations of speech sounds occur together, even if they only hear the sounds several times in a few minutes.

By 9 months, infants seem to recognize the phontactic and stress patterns of their native speech, and they use this information in segmenting words, and they remember words that occur in speech. Now they are sensitive to the acoustic cues that mark phrase (not just clause) boundaries.

We'll come back to the non-native speech issue later, but 10-12 months seems to be an important time for native language learning.

By 10.5 months, infants use allophonic information to segment words, and English learning infants can now segment words that have initial unstressed syllables (e.g., guitar)

By 12 months they seem to be able to use several segmentation cues together.

Do children use acoustic cues in speech like adults?

Booed or Boot?

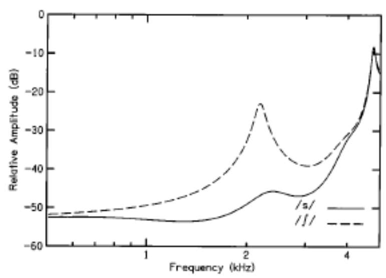
Cab or Cap?

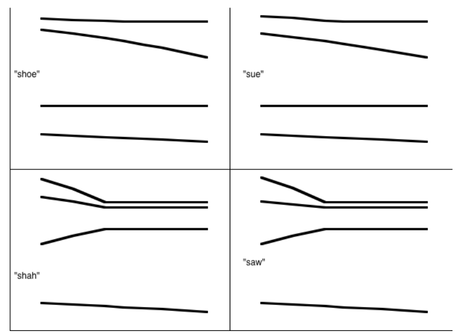
- VC transition
- Vowel duration
- Burst frequency
- Etc.Children’s use of formant transitions and static spectral information in speech

By the end of infancy, children seem to know a lot about speech. They know words; they produce short sentences; they are able to segment words from running speech.

A question that has received some attention is whether children are using the same information that adults use when they identify speech. For example, when we distinguish the word “booed” from the word “boot”, or the word “cap” from “cab”, we can use the duration of the transition between the vowel and the final consonant, the duration of the vowel, the frequency of the burst that accompanies release of the final consonant, as well as other acoustic features. Adults may depend on one of these features under most conditions, but if that feature is had to hear for some reason (e.g., if it is masked by competing sound), adults can switch to another cue. Moreover, they may use on cue more in some phonetic contexts than in others.

Children’s use of formant transitions and static spectral information in speech





Nittrouer took 4 syllables: sue, shoe, saw and shah. These syllables are distinguished in two ways. The spectrum of the noise accompanying the fricative is higher for /sh/ than /s/. This is illustrated in the graph on the left of the slide. In addition, the formant transitions into the vowel are different. This is illustrated on the right of the slide. The onset frequencies of F2 and F3 are closer together for /s/ than for /sh/, for both /a/ and /u/ contexts. Nittrouer asked what children would do if you paired the formant structure for /sh/ with the noise for /s/, and the formant structure for /s/ with the noise for /sh/. If a listener isn’t paying much attention to the noise spectrum, then he or she might judge the sound only on the basis of the formant structure. Similarly, if the listener is only paying attention to the noise spectrum, he or she might judge only on the basis of the noise spectrum.

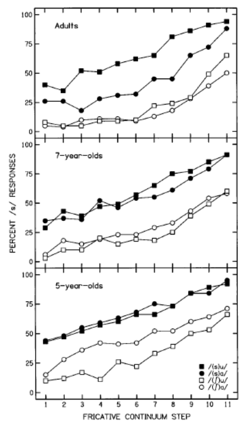
Nittrouer, S. (2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *J. Acoust. Soc. Am., 115*(4), 1777-1790.

Nittrouer, S. (2005). Age-related differences in weighting and masking of two cues to word-final stop voicing in noise. *J. Acoust. Soc. Am., 118*(2), 1072-1088.

Nittrouer, S., & Crowther, C. S. (1998). Examining the role of auditory sensitivity in the developmental weighting shift. *J. Speech Lang. Hear. Res., 41*(4), 809-818.

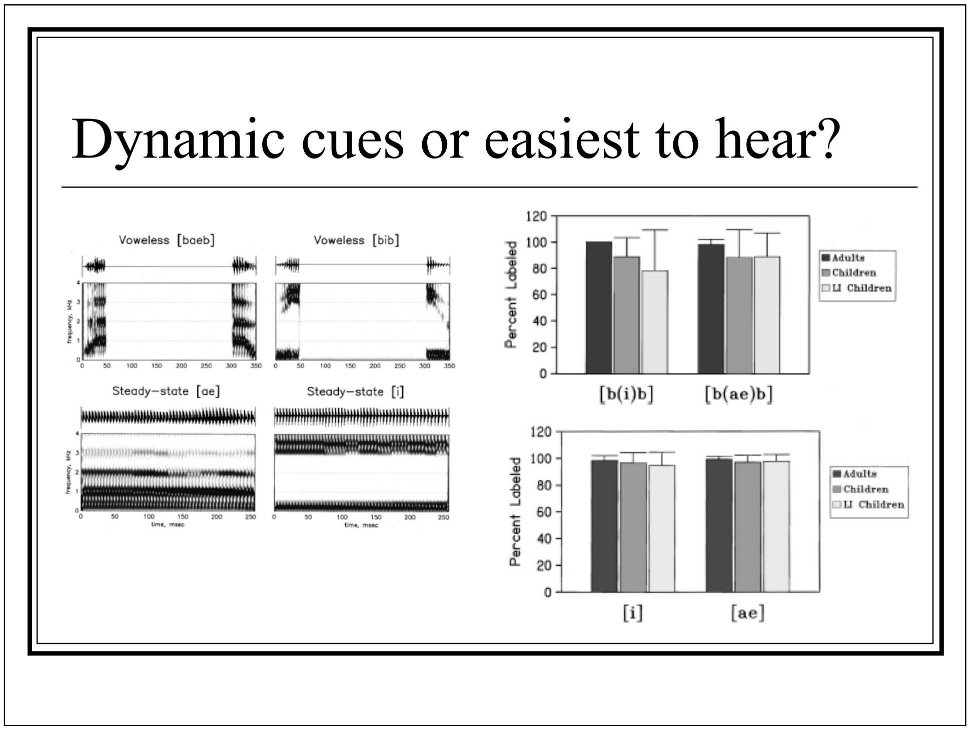
Nittrouer, S., Crowther, C. S., & Miller, M. E. (1998). The relative weighting of acoustic properties in the perception of s+stop clusters by

Children's use of formant transitions and static spectral information in speech



These are the results of one such experiment with adults, 7 and 5 year-old children. Where the results for syllables starting in /s/ noise are shown in filled symbols, and for /sh/ noise are shown in unfilled symbols. Squares are for /u/ vowel and the circles are for /a/ vowels. The x-axis is the peak frequency of the frication noise-- as it shifts from high to low (1 is high and 11 is low), the listener is expected to say /s/ more often. If nothing else made a difference, then as the fricative continuum step increased, /s/ judgments should increase, and the lines would all fall on top of each other. If the person is also paying attention to the formant structure then the syllables with /sh/ formant structure will have curves that fall lower than the ones with /s/ formant structure. The slope of the curve tells us how much the listener is using the noise spectrum and the distance between the filled and unfilled symbols tells us how much the listener is paying attention to the formant structure. Adults seem to use both piece of information, but they use the formant structure more in /u/ context than in /a/ context.

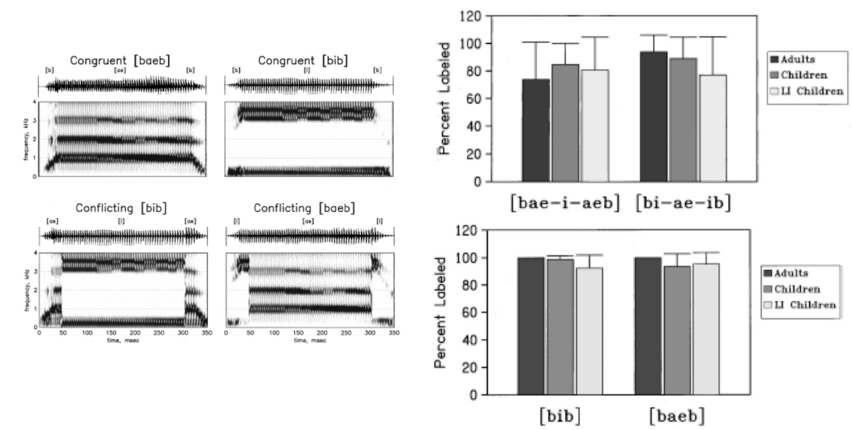
Now consider the 5-year olds. First, the slopes of their functions are shallower than those of adults. That suggests that they are not paying so much attention to the noise spectrum, but their results for the /s/ formant structure and /sh/ formant structure (filled and unfilled squares) are separated pretty much like adults, at least in the /u/ context. The results in the /a/ context are closer together, also like the adults, though the curves both fall up higher. The results for the 7-year-olds are a little odd, in that they now look like they used the formant structure just the same for /a/ and /u/ context, unlike either adults or 5-year-olds.



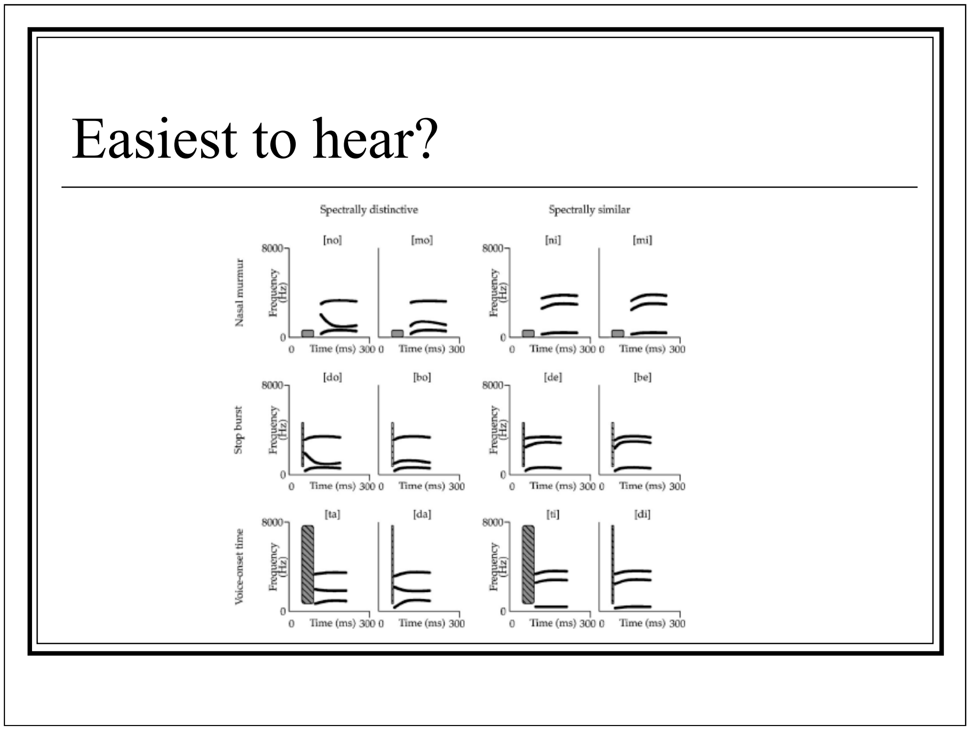
Sussman (2001) argued that kids use the transition information not because it is dynamic or defines the shape of the syllable, but because they can hear the transitions better than they can hear the spectral details. Although kids at this age have mature frequency resolution, it might still be that they have trouble extracting some information from complex sounds. Sussman created stimuli like the ones shown in the spectrograms on the left of this slide. They are the syllables “beeb” and “bab” with the whole steady-state part of the vowel cut out-- the transitions are there, but nothing else is. Performance on the identification of the steady state vowels alone is also examined. If Nittrouer’s idea about dynamic cues or syllable shape is correct, then the kids should do fine in these “voweless” syllables. The results for adults and children are shown on the right. (There are results shown for kids with language impairments “LI children” that I won’t discuss). Notice that for the vowels alone, the children do as well as the adults, but that they don’t do as well as adults for voweless syllables-- the transitions alone don’t seem to do it for them.

Sussman, J. E. (2001). Vowel perception by adults and children with normal language and specific language impairment: Based on steady states or transitions? *J. Acoust. Soc. Am., 109*(3), 1173-1180.

Dynamic cues or easiest to hear?



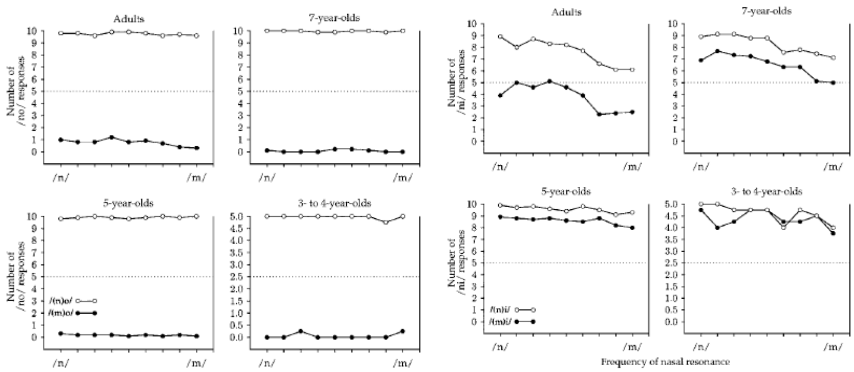
In a second set of conditions, Sussman tested the conditions in which either the correct (congruent) or wrong (conflicting) vowel was added to the voweless syllables, "a" was added to the syllable. The question is, will the listener pay attention to the transition, or to the steady state formants of the vowel? The children and adults perform similarly in both conditions, suggesting that children and adults are attending to similar acoustic information in the syllables. Sussman argues that in fact, when the transitions aren't that easy to use (when they are presented alone), the children don't use them as well as the adults do.



Finally, Mayo and Turk tried to settle this question-- do children depend more on the most salient or “easy to hear” acoustic cues. They chose stimuli with the stylized spectrograms shown in this figure. They judged that the pairs of syllables on the left were spectrally distinctive-- at least with your eyeballs it looks like it would be easy to see that the formants were different in frequency, while the syllables on the right have very similar formant structure, Mayo and Turk hypothesized that if kids used information that they could hear easily, they would have more trouble with the spectrally similar contrasts than with the spectrally dissimilar contrasts.The other cue that was manipulated is represented by the boxes at the beginning of the syllable. For /m/ v. /n/ it is the nasal murmur (frequencies of F2 and F3 in the initial part of the syllable); for /t/ v. /d/ it was voice onset time. And for /d/ v. /b/ it was the spectral shape: for /d/, the format peaks rise in intensity from F1 to F2 to F3, while for /b/, they fall. So the experiment is: tell me whether this is a /b/ or /d/ as I gradually change the spectral shape from one to the other, but in one condition the transitions belong to /b/ and in the other the transitions belong to /d/.

Mayo, C., & Turk, A. (2005). The influence of spectral distinctiveness on acoustic cue weighting in children's and adults' speech perception. *J. Acoust. Soc. Am., 118*(3 Pt 1), 1730-1741.

Sometimes spectrally dissimilar is better

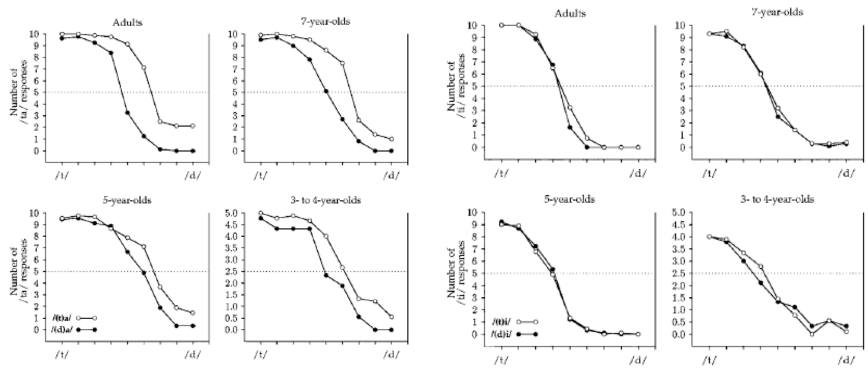


Some of Mayo and Turk’s results are consistent with the idea that kids need easy to hear distinctions. For example, here are the results for the /n/-/m/ distinction. The open symbols represent the syllables that had /n/ transitions and the filled symbols are for syllables that had the /m/ transitions. The graphs on the left are for /mo/ v /no/ and the ones on the right are for /ni/ v /mi/.. /mo/-/no/ is distinctive; /mi/-/ni/ is less distinctive. Each graph is for a different age group, and plots the proportion of times the listener said “no”, as the sound was gradually changed from /no/ to /mo/-- but remember that even with these changes, the syllables always had either /m/ or /n/ transitions.. As in Nittrouer’s studies, the separation between the curves in these graphs tells us how much people depended on the transitions, and the slope of the curve tells us how much they depended on the other cue-- here the nasal murmur. For the spectrally dissimilar (distinct) mo-no contrast, everyone always says “no” when the syllable has /n/ transitions, and “mo” when the syllables had /m/ transitions.Nobody really cares about the nasal murmur.

On the right you see what happens for /ni/ v /mi/. Notice that the curves are closer together for the adults-- as if they can’t figure out which one is which, and as the syllables become more “mi”-like, the adults tend to say “ni” less often. That is, they start to pay attention to the nasal murmur. The results for the kids are different. The younger the child, the closer together the curves are-- the children have more trouble distinguishing these spectrally similar sounds. In addition, for younger children there is less of a tendency to start saying “mi” as the syllable becomes more “mi”-like. They don’t start to use the other cue, even though the transitions are hard to distinguish.

So these results are consistent with the idea that children have trouble hearing some acoustic cues in speech. This was true in Mayo and Turk’s results for /m/ and /n/ and for /d/ and /b/, but interestingly, it wasn’t true for the /d/-/t/ distinction.

Sometimes children are like adults



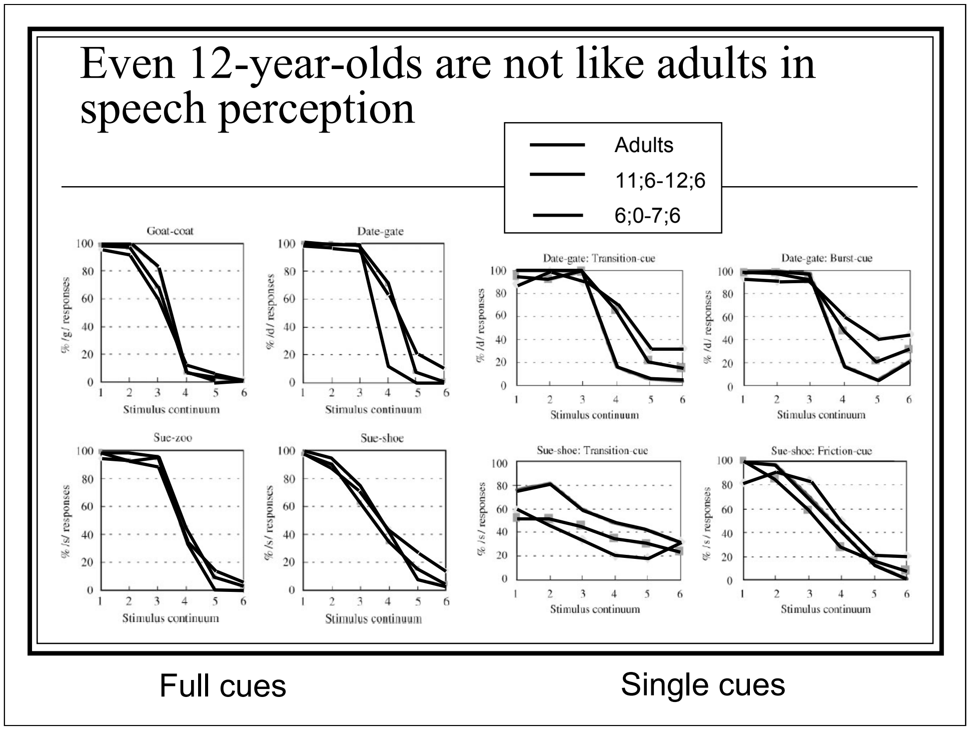
The format here is the same as in the last slide. With the spectrally dissimilar contrast on the left, and the spectrally similar contrast on the right. Each graph plots the percent of /d/ responses as the frequency of noise burst is shifted from /d/-like to /t/-like. . The unfilled symbols are for /d/ transition syllables and the filled symbols are for /t/ transition syllables. The two curves are separated for the adults, but less separated for the children, indicating that they are less able to use the spectral dissimilarity to distinguish these syllables, although all age groups say /d/ less often as VOT is changed. Note that this pattern is the opposite of what Nittrouer has reported with fricative and other contrasts. When the syllables are spectrally similar (right graph), everybody ignores the transitions and pays attention to the VOT.

Are 6-year-olds like adults in speech perception?

Phonemic contrast	Minimal pair	Acoustic cues varied
/γ/-/κ/ Stop voicing	Goat–Coat *	Voice onset time F_1 transition
/δ/-/γ/ Stop place	Date–Gate *	Burst frequency F_2/F_3 transitions
/σ/-/ζ/ Fricative voicing	Sue–Zoo	Friction duration Voicing amplitude
/σ/-/Σ/ Fricative place	Sue–Shoe *	Friction frequency F_2 transition

Finally, Hazan and Barrett examined the performance of school aged children in speech discrimination in two conditions. They asked kids to label exemplars of the words shown in this table. Several exemplars of each word pair were constructed so that they varied acoustically from a clear example of one word, through some more ambiguous tokens, to a clear example of the other word. For each word pair two different acoustic cues were manipulated,

In one condition, they left two acoustic cues to the word’s identity in the word. In the other condition, they retained only one of these cues. So the listener might have to switch the cue she or she is using in the single-cue case.



The results are shown here. In each panel is plotted the proportion of time that the listener labeled the word as the first item of the pair. So in the top left panel, is plotted the proportion of exemplars labeled goat for each of the exemplars along the acoustic continuum. And you can see that as you move along the continuum , listeners are less and less likely to label the exemplar goat (and of course more and more likely to label the exemplar coat). In each panel the adult's responses are shown in black, the 6-8 years olds are showing red and the 9-13 years olds are shown in blue. Notice that for each of the four word pairs in the left hand graphs, the children's identification function is just a little shallower than the adults. A small difference, but pretty consistent across contrasts, indicating that the children are somewhat less consistent than the adults in the way that they label these sounds.. However, look what happens when only one of the two cues is available in the stimulus, In the right hand figures are shown the results of the date versus gate comparison first when only a transition cue is included, and then when only the burst information is available. Now the children have much shallower identification functions than the adults-- they could use the two cues together pretty well, but now when they are asked to depend on just one they have trouble.

The same pattern is seen for the sue-shoe contrast in the bottom right figures.

One way to interpret this is that the children are less flexible in the way they process speech. adults can switch from one cue to the other if they have to, and it's clear that under different listening conditions in noise or fast speaking rates, for example, they would normally need to be able to do that. Children on the other hand don't seem to be able to switch cues as needed.

So the big conclusion is that it takes a long time for speech perception to become as sensitive, complete, and flexible as what is seen in adults.

Conclusions

- Early in infancy, auditory immaturity may limit speech perception.
- Even though infants and children can discriminate speech sounds, they use different cues than adults do.
- Even school-age children have difficulty switching from one cue to another in speech discrimination when listening conditions change.

Overview of effects of experience on speech perception

Nature v. Nurture?

It's always both.

We know experience hearing speech changes speech perception.

- Children learn to speak the language they hear around them.
- Adults have difficulty discriminating and producing some contrasts in nonnative languages.

How could experience with speech affect development of speech perception?

- Hardly at all
- Attunement
- Maintenance
- Facilitation
- Induction

We might ask how specifically experience has its influence on the development of speech perception.

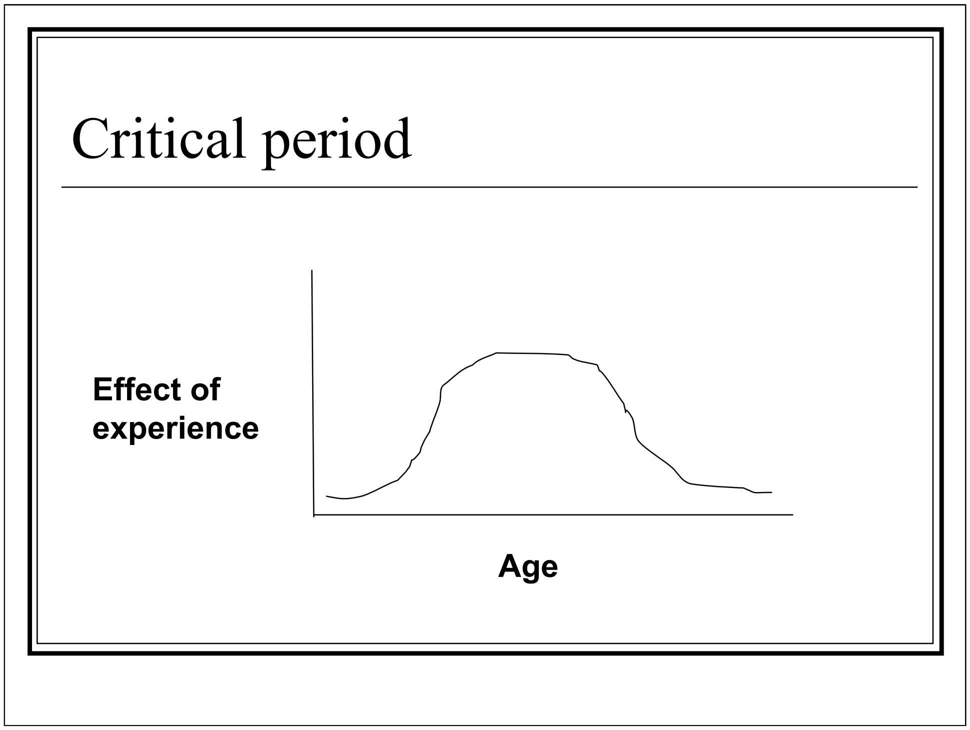
Of course, it could be that experience hearing speech doesn't matter at al It could be that there are some speech contrasts that occur in all languages in exactly the same way, and people will be able to discriminat them not matter what they hear.

Another possibility is that the ability to discriminate a contrast is there whe we are born, but that it takes some experience with hearing that contrast t sharpen up or improve or slightly change our discrimination. I won't talk about this today, but we know it occurs. For example, the VOT boundary between voiced and unvoiced consonants is often somewhat different in different languages, and it appears that experience just shifts the bounda over a bit.

Or it could be that experience is that we are able to discriminate some contrast, but that experience with that contrast is necessary to maintain that ability. For example, maybe we are all able to discriminate a certain contrast when we are born, but that unless we hear that contrast, we lose that ability.

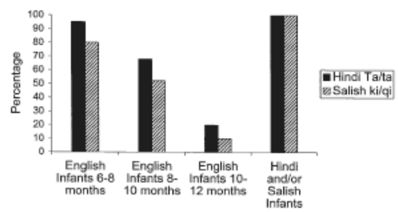
Another possibility is that experience hearing speech facilitates or makes easier, the ability to discriminate some sounds. In other words, if you don' hear the contrast, you will eventually be able to discriminate it , but if you hear it a lot, you will be able to do so sooner.

Finally, it is possible that you will never develop the ability to discriminate some contrast unless you hear it.



The other thing that we need to keep in mind is the timing of the experience. It may be that there is a critical period during which experience hearing speech must occur in order for it to be effective. And the critical period could conceivable be difference for different speech sounds.

Maintenance role of experience



The results of a study by Werker and Tees really demonstrate the role of experience in maintaining perception of a contrast. They had English learning infants and adults discriminate contrasts that occur in Hindi or in a native American, Salish, language that do not occur in English and that it is unlikely that English learners would have heard. They used the conditioned head-turn technique to test discrimination. What they found was that 6-8 month-old infants responded to a change in a nonnative syllable a high proportion of the time. 8-10 month olds responded a little less often, and 10-12 month olds hardly responded at all. Between 6 and 12 months, infants stopped making this discrimination. However, infants who were learning Hindi or Salish were able to discriminate the contrasts 100% of the time even at 12 months. English speaking adults performed about as well as the 10-12 month olds.

Werker, J., & Tees, R. (1993). The organization and reorganization of human speech perception. *Ann. Rev. Neurosci., 15*, 377-402.

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Dev., 52*, 349-355.

Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Can. J. Psych., 37*(2), 278-286.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.*, 7, 49-63.

Later research showed that

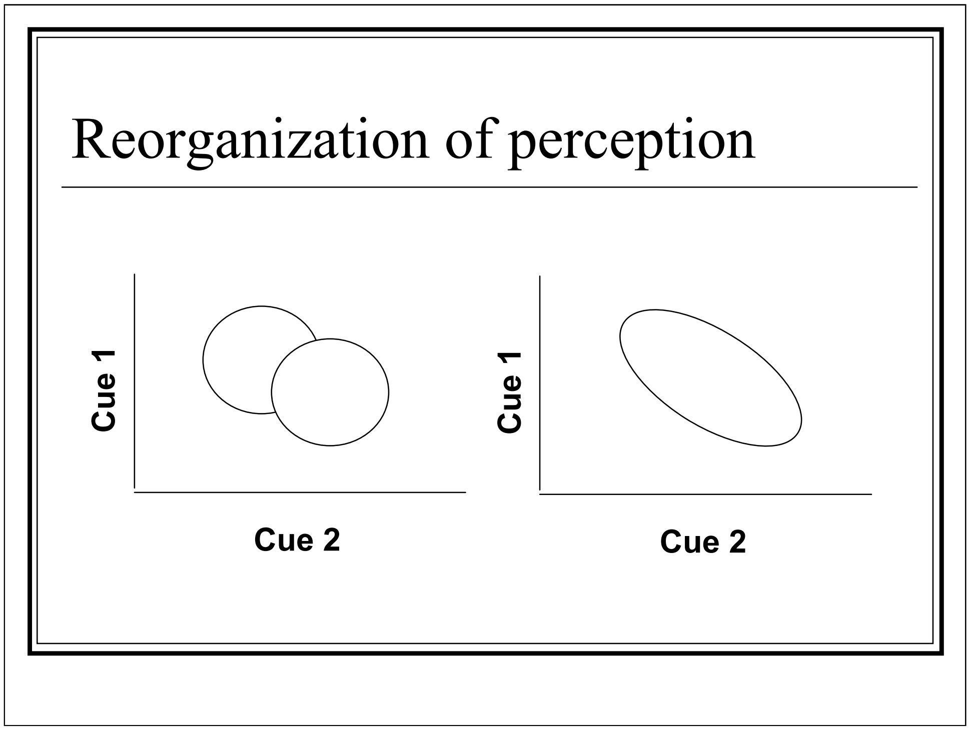
- Sometimes 12-month-olds lose the nonnative discrimination, but adults are able to discriminate.
- Sometimes the nonnative discrimination seems unaffected by experience.
- The role of experience seems to be to reorganize perception in a way that makes it hard to discriminate nonnative contrasts.
- With the right kind of experience, adult listeners can learn to make nonnative discriminations.

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 289-304). Dordrecht: Kluwer Academic Publishers.

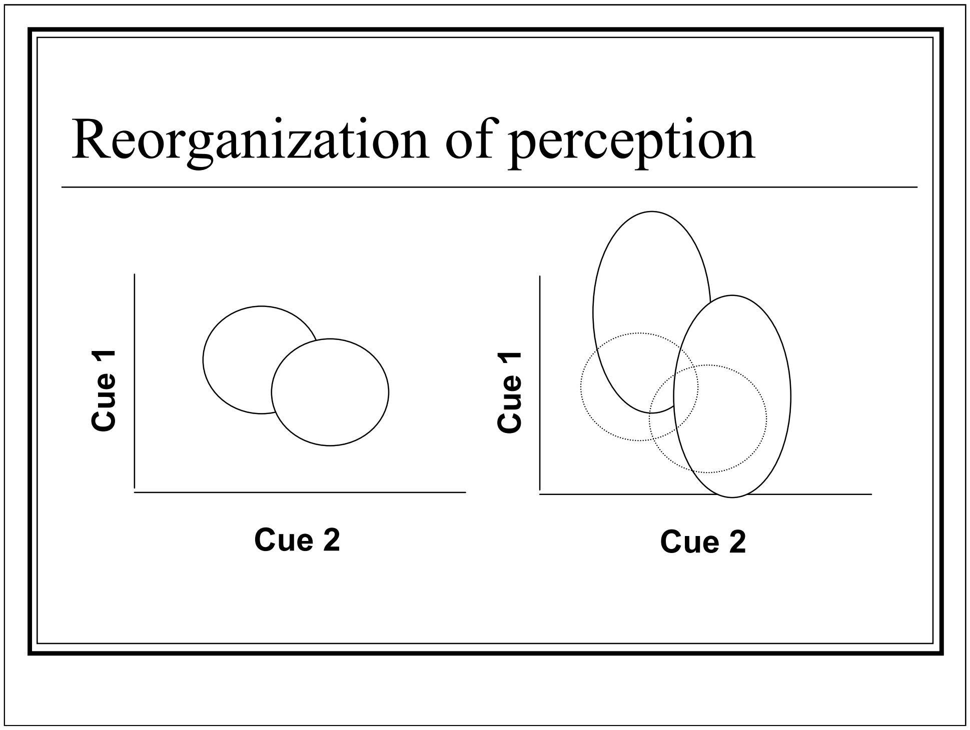
Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age [see comments]. *Science, 255*(5044), 606-608.

Polka, L., & Bohn, O.-S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *J. Acoust. Soc. Am., 100*(1), 577-592.

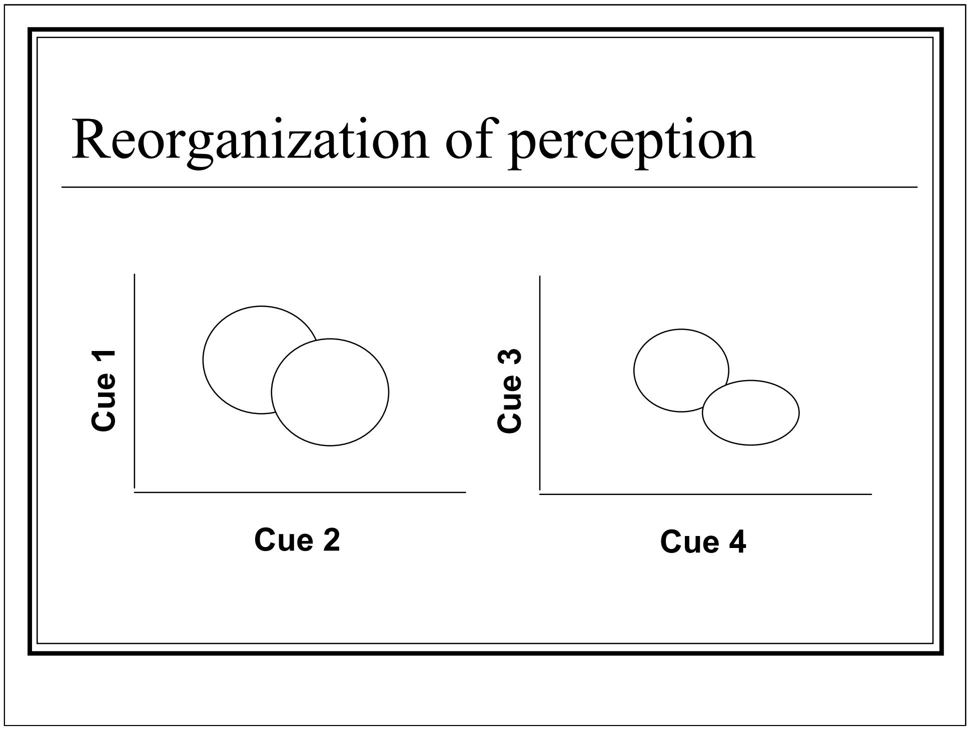
Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *J. Exp. Psychol. [Hum. Percept.], 20*(2), 421-435.



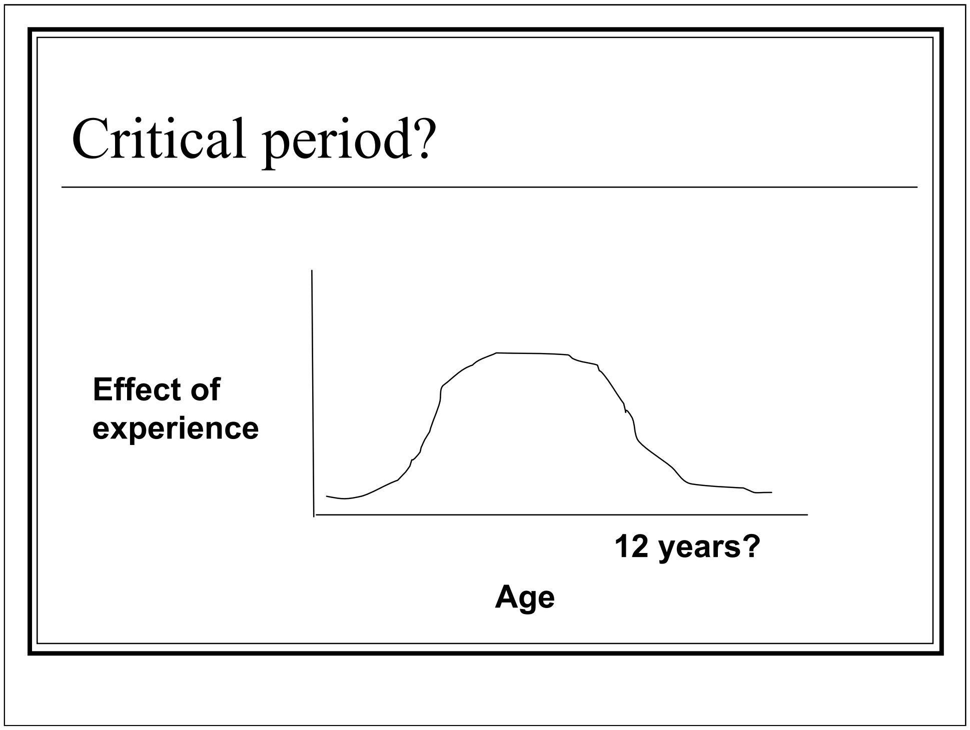
Imagine that we have a perceptual space defined by some combination of acoustic cues. Young infants may be able to discriminate diffeent categories of speech sounds like a native speaker, but if infants never hear this distinction made because in the native language the sounds in these two categories are consider the same category, then their experience will lead to the merging of the two categories, and now sounds they used to be able to discriminate are no longer discriminable. On example of this is the /r/-/l/ distinction for Japanese speakers.



Another possibility is that the original categories get modified by experience-- the categories are now defined somewhat differently. But if the boundary between the new categories divides some of the exemplars in the old categories overlap some with the new ones, people might still be able to make the discrimination at some level. An example of this is the French lateral fricative voicing contrast for English speakers (which I couldn’t possibly say); a similar difference occurs between /s/ and /z/ in English.



Another possibility is that the original categories are defined by one set of acoustic cues thati isn't even used in the native language. Nonnative speakers listening to such sounds may not even think they are speech. In these cases the nonnative speakers may retain the ability to discriminate the original categories, because their experience hasn’t changed the organization of those sounds. An example is with click sounds used in Zulu languages for English speakers.



On the basis of studies of children who for unfortunate reasons do not learn speech as infants, people have drawn the general conclusion, that humans are able to learn language until they are 12 years old or so. But no studies have really addressed whether speech is different from language in general, nor have any studies looked in detail at how children learn "new" speech contrasts compared to adults and how that might vary by age. The general belief is that kids can learn new speech sounds more easily than adults-- maybe even teenagers learn better than adults, but the data outlining the details are not available.

Effects of experience on development of speech perception

- Experience with native speech leads to a reorganization of perception.
- The way that the reorganization will affect performance depends upon the relationship between the original speech categories and the native speech categories.
- Examples of no influence, attunement, maintenance and facilitation of speech perception by experience hearing speech have been identified.