

Native listeners

ANNE CUTLER

Max Planck Institute for Psycholinguistics and University of Nijmegen,
PO Box 310, 6500 AH Nijmegen, The Netherlands. E-mail:
anne.cutler@mpi.nl

Becoming a native listener is the necessary precursor to becoming a native speaker. Babies in the first year of life undertake a remarkable amount of work; by the time they begin to speak, they have perceptually mastered the phonological repertoire and phoneme co-occurrence probabilities of the native language, and they can locate familiar word-forms in novel continuous-speech contexts. The skills acquired at this early stage form a necessary part of adult listening. However, the same native listening skills also underlie problems in listening to a late-acquired non-native language, accounting for why in such a case listening (an innate ability) is sometimes paradoxically more difficult than, for instance, reading (a learned ability).

Introduction

When this material was presented as a lecture at the Academia Europaea meeting in Rotterdam in the summer of 2001, the members of the audience were effectively taking part in the kind of experiment described below. Some were native speakers of English; most were not. The lecture was presented in English. The topic is listening to spoken language and, in particular, why listening to non-native language can be so much harder than listening to the native language. It is to be hoped that the non-native majority of the audience nevertheless found the lecture adequately comprehensible.

Difficulty in listening to non-native language is somewhat paradoxical. Listening is a natural ability, and exposure to spoken language in infancy results naturally in acquisition of language skills. Reading is an unnatural ability; reading skills must be explicitly taught. Yet highly proficient users of non-native language frequently observe that reading in the second language can be easier than listening. Even when the orthography of the second language does not match that of the first (Japanese adult learners of English, for example), written text can be perceived as easier to understand than, say, a congress presentation. It is particularly

noticeable that listening skills are less flexible with a non-native language: whereas, in our native language, we can cope seemingly effortlessly with previously unheard voices, severe bandpass restrictions or substantial background noise, non-native listening can suddenly become much harder with unfamiliar speakers or under difficult listening conditions.

Part of the reason for this asymmetry is the very naturalness of language listening skills: the fact that they underlie infant language acquisition. Although it is not until around their first birthday that infants give overt evidence of linguistic communication abilities, they are in fact hard at work at developing the foundations of these abilities throughout the first year of life. Some relevant research is summarized in the following section.

Infant listening

A baby under the age of one seems to spend most waking time lying around gurgling or wriggling; the development of listening skills is not immediately reflected in behaviour. However, behavioural experimental techniques developed over the past couple of decades have enabled researchers to observe the gradual refinement of phonetic perception, in particular across the second half of the first year of life.

Six-month-olds, for example, can learn to respond to speech sound contrasts. They might hear a series of syllables such as *ta, ta, ta ...*, which at some point switches to *da, da, da ...*; if, on hearing the change in the input, they turn their head in a particular direction, a reward appears (as it might be, an animated puppet is illuminated for a short time). Correct head turns when the stimulus changes indicate that the baby has noticed the change; failure to respond suggests that the change has not been noticed, i.e. that a particular phonetic contrast is indistinguishable to the child. Babies of six months of age perform this task as well with contrasts from non-native languages (i.e. contrasts which are not distinctive in the language to which they have been exposed) as with contrasts that are distinctive in the environmental language. The left panel of Figure 1 shows a representative observation.

At ten months of age, however, babies cannot any longer reliably make the non-native distinctions, although they still successfully distinguish the native contrasts (middle panel of Figure 1). Their performance at 10 months in fact resembles adult performance – adult listeners (right panel of Figure 1) can only reliably discriminate native contrasts, and show no more than a 50% chance of distinguishing non-native contrasts. (Note that adult listeners need no animated puppets as reward; they are willing to register detection of a speech sound change by pressing a response key. In other respects, however, the infant and adult experiments are comparable.)

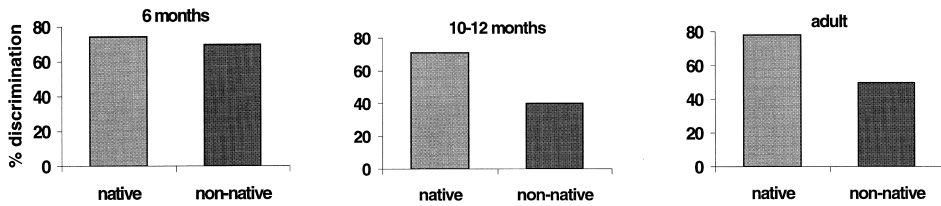


Figure 1. Percentage of responses showing discrimination, (a) in 6-month-old infant listeners, (b) in infants of 10–12 months, (c) in adult listeners (results of Werker and Lalonde¹⁴).

By around ten months of age, therefore, babies' phonetic perception has become attuned to the contrasts that are relevant for the environmental language; other contrasts are disregarded. Further, babies of this age are familiar with the relative probabilities with which phoneme sequences occur in the native language. This has been demonstrated by research making use of preference tasks, i.e. tasks that not only measure infants' ability to discriminate between two inputs, but also determine whether infants exhibit a preference for one of the two. For instance, if infants can control the presentation of an auditory stimulus by maintenance of a headturn, a preference will be displayed when they choose to listen longer to the one input than to the other.

Such methods were used for instance in a study by Jusczyk *et al.*¹ They presented English-learning infants with lists of nonsense words such as *riss* and *goove*. Although neither of these is a real English word, *riss* is made up of highly probable phonetic sequences (that is, many English words contain *ri-* and *-iss* sound sequences: *rib*, *rich*, *rinse*, *hiss*, *miss*, *fist*, *crisp* etc). *Goove* is much less probable, since there are relatively few English words containing *goo-* or *-oove* sequences (such as *goose*, *move*, *prove*). At eight to ten months (though not yet at six months), Jusczyk *et al.*'s infant subjects listened longer to the lists made up of probable sequences than to the lists made up of less probable sequences. This suggests that the refinement of listening skills after the age of six months includes an appreciation of phonetic sequencing likelihoods in the input language. Another study by Jusczyk *et al.*,² also with the listening preference technique, showed that infants also in fact recognize word structure probabilities of the native language – their nine-month-old subjects listened longer to lists of English words with stress on the initial syllable (*former*, *butter*, *curdle* etc.) than to lists with stress on the second syllable (*inform*, *abut*, *occur* etc). The most common kind of word in the English vocabulary is a bi-syllable with initial stress; apparently this distributional preponderance is obvious even to nine-month-olds. Again, however, six-month-olds did not display such a preference. Familiarity with word structure probabilities is thus also acquired between six and nine months.

The characteristics of the input at this crucial period of life were examined by Joost van de Weijer in a PhD project in our laboratory.³ He analysed a corpus consisting of the entire input to a baby from the age of six to the age of nine months, collected via a tape recorder that accompanied the baby during all her waking hours. One of the most noteworthy findings of his study was that nearly all the language that the baby heard consisted of connected speech; isolated words made up less than 9% of the corpus. Connected speech is notoriously continuous: although to adult ears speech consists of a string of separable words, the acoustic reality is that words run on from one another without a break, and there are few robust and reliable signals of boundaries between words. Thus, babies are able to construct accurate representations of word structure probabilities even when presented with an input in which the individual words are not clearly demarcated. This in turn suggests that babies are capable of extracting individual words from continuous speech.

This has indeed been demonstrated in the infant speech perception laboratory. Juszyk and Aslin⁴ adapted the listening preference procedure by adding an initial familiarization phase. For instance, they presented babies with multiple repetitions of a pair of words – and not necessarily words that might reasonably have been known to infants. Thus, in the original experiment of Juszyk and Aslin two of the words were *feet*, *bike*; a later experiment used words such as *hamlet*, *kingdom*. In the second phase of the experiment, the babies were presented with short passages, and their listening preferences were assessed. Two of the passages contained multiple repetitions of the familiarized words; the other two passages did not contain the familiarized words. Babies of six months showed no preference for one type of passage over the other, but eight-month-old babies did show a preference: they listened significantly longer to the passages containing the familiarized words (see Figure 2). This indicates that they had recognized the words, which they had originally heard in isolation, in their recurrence in the continuous-speech contexts.

Babies of nine months of age can, indeed, even perform this feat with input in a non-native language (at least in a language which in many respects resembles the native language). Houston *et al.*⁵ found that Dutch-learning nine-month-olds and nine-month-olds from American English-speaking families showed almost exactly equivalent preferences for Dutch passages containing words with which they had just been familiarized (Dutch words such as *pendel*, *karper*), in comparison to totally unfamiliar passages (see Figure 3).

At the end of the first year of life, therefore, babies have achieved a great deal. They have built up knowledge of the phoneme repertoire of the native language and targeted their phonetic sensitivity at the contrasts relevant for native phoneme discrimination. They have learned the relative probabilities of phoneme sequences and word structures. And they have learned to recognize words heard in the

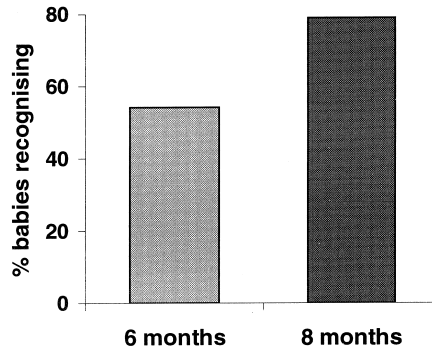


Figure 2. Percentage of infant listeners (6 months versus 8 months) who recognized familiarized words in a continuous-speech context (results of Jusczyk & Aslin⁴).

context of continuous speech. Their work has put them in a position to begin the process of building up a vocabulary and using it for spoken communication. A baby of ten months or so may be regarded as, phonetically speaking, a mature listener; we shall see below, however, that it may be as accurate to regard an adult listener as, phonetically speaking, effectively an infant.

Native listening

One major difference between infant and adult listeners is that the adult listener's vocabulary is in place. Adults know tens, perhaps hundreds, of thousands of words. However, these many words are not necessarily easily discriminable. For instance, they are made up of relatively few phonemes. On average, languages

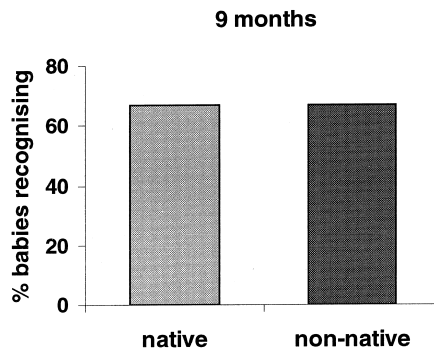


Figure 3. Percentage of Dutch and American 9-month-olds who recognized familiarized Dutch words in a Dutch continuous-speech context (results of Houston *et al.*⁵).

have inventories of 30 or so phonemes; in the largest representative sample of phoneme inventories,⁶ the median and mode of the sample both lie below 30. (English, with over 40 phonemes, thus counts as a phoneme-rich language. Dutch, with 35, is somewhat closer to the mean.)

In addition, languages prefer short words to long ones; and severe co-occurrence restrictions obtain (for instance, strings of consonants without vowels would be unpronounceable). The inevitable consequence of constructing a large vocabulary with many short words out of a very small set of phonemes, which may furthermore not occur in all conceivable sequences, is that words resemble one another, and longer words tend to contain shorter ones as accidental embeddings. Add to this the problem of the continuity of natural speech, in which words abut one another without cues to signal the boundaries between them, and it is clear that the recognition of spoken words in natural conversation presents the adult listener with a formidable task.

How listeners nevertheless perform the remarkable feat of recognizing spoken language without noticeable delay or error has been the subject of a considerable research effort in the last two decades. We now know that spoken-language input automatically activates in the mind of the listener all the words that it supports – including words that are not part of the speaker’s intention. That is, words that are only spuriously present in an utterance via accidental embedding may also be activated; an utterance of the English word *startle* may also activate *star*, *tar*, *art*, *start* and *tart*, an utterance of the Dutch word *luisteraar* may also activate *lui*, *luis*, *te*, *raar* and *aar*. Likewise, words may be activated when they are accidentally present via the juxtaposition of two other words – English *start* in *best article*, Dutch *altaar* in *meestal taart*, German *gering* in *Gebirge ringsum*. Even partial information about a word may produce activation – thus the beginning *sta-* may activate *star* and *start* and *startle* and *stark* and *starling* and *starving*, all at once. In short, many more words are activated than should actually be recognized; the listener’s task is to recognize those words that are indeed intended by the speaker, and jettison the rest.

This too occurs automatically, via a process of competition between the simultaneously active word candidates. The competition process is won by the words that are best supported by the signal (*startle* supports *startle* to a greater extent than *tar* etc), or which provide the best total account of the input (in *star vehicle*, for instance, the input temporarily supports *starve* to a greater extent than *star*, but the [v] is eventually claimed by *vehicle* as a result of which *starve* loses its extra modicum of support, and *star* wins out).

Psycholinguists cannot observe the processes of spoken-word recognition directly, but they have developed many ingenious methods for studying lexical activation and inter-word competition indirectly in the laboratory. One such method involves recording a listener’s eye movements during the performance of

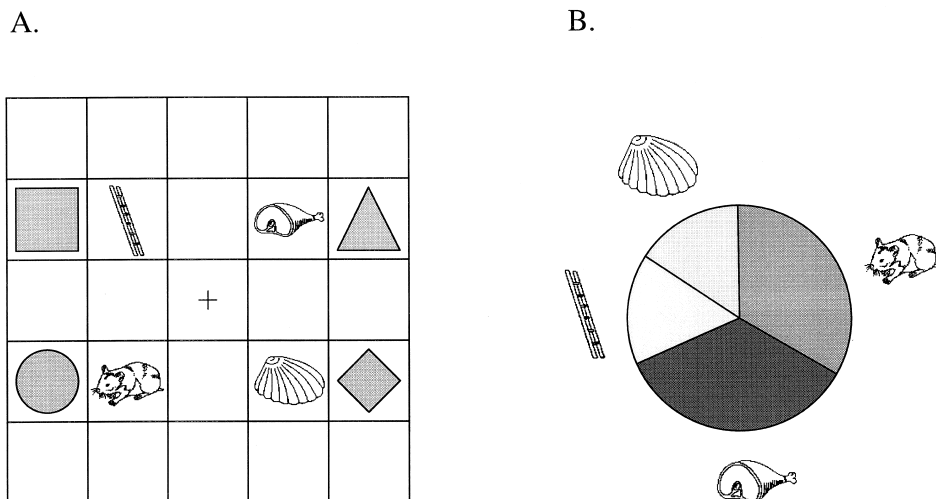


Figure 4. (a) Stimulus display for eyetracking experiment (Dutch listeners, Dutch input); (b) Percentage of eye fixations at the point that listeners have heard *click on the ha-* (*pak de ha-*; results from Salverda⁷).

a simple task to spoken instructions. A small harness, placed on the listener's head, supports a camera that monitors the direction of gaze. In front of the listener is a display such as that in Figure 4(a), containing depictions of several simple objects and several geometric figures. The instructions the listener hears may then be *Click on the hamster and move it underneath the triangle*.

The depicted display contains, besides a *hamster*, also a drawing of a *ham*. At the point at which the listener hears *click on the ha-*, it is unclear to which of these two objects the utterance will refer. Typically, however, listeners do not wait until the ambiguity has been resolved, but – as illustrated in Figure 4(b), which shows results from the dissertation research of Anne-Pier Salverda⁷ – they look at one or other of the two partially supported candidates. Neither of the two is at this point clearly preferred, but both are equally available, and both are also clearly preferred over the other available objects, in this case the ladder and the shell.

Another laboratory method with which one can clearly see effects of activation and competition is the so-called word-spotting task. Listeners hear a series of short nonsense strings and, as the name of the technique suggests, their task is to spot real words in the input. Some of the non-words indeed begin or end with real words; for instance, in the sequence *glervo deprull sleepnah skelth foomrock hezling mintowf raftaysh* there are four words to be found: *sleep*, *rock*, *mint*, *raft*. The listener does not know in advance how many words might appear or what they are; when a word is spotted, the listener signals detection by pressing a

response button, and then reports the word. Speed of response and accuracy of detection can both be measured.

The word-spotting technique is supposed to provide a microcosmic version of word recognition in continuous speech; embedded words are provided with a minimal context, and word and context are, just like adjacent words in normal speech, smoothly continuous. The relative difficulty of different contexts for detection of a particular word can be compared. For example, the detection of *mint* is easier in *mintowf* (as in the example above) than in the hardly different sequence *mintayf*.⁸ The reason for this is that the two contexts *-owf* and *-ayf* differ in the amount of competition they call into being. In both cases the appended context begins with a vowel, and this vowel combines readily with the final consonant of *mint* to activate potential words beginning with that consonant-vowel sequence (just as *ha-* effectively activates both *ham* and *hamster*). The difference is that there are many English words beginning with [t] plus the vowel in *-ayf* (*table, take, tail, tape, tame, taste* etc), but very few beginning with [t] plus the vowel in *-owf* (*towel, town, tower, tout*). Thus, the context *-ayf* causes more competition for the [t] of *mint* than the *-owf* context does; detection of *mint* is retarded by this competition, and thus takes longer in the *-ayf* context than in the *-owf* context.

Listeners do have at their disposal a number of effective ways of modulating the competition process. For instance, they can be effectively assisted by phoneme sequence restrictions and probabilities (which, as we saw above, are apprehended even in infancy). Thus, the word *rock* can be detected more readily in *foomrock* than in *fooprock*;⁹ the context *foom-* cannot easily combine with *rock*, because no English word contains a syllable beginning or ending with [mr]. In contrast, *pr-* is the beginning of many English words (*prod, pray, proud, preen* etc), and the context *foop-* might thus activate potentially injurious competition. But activation of many candidate words simultaneously is not necessarily harmful – competition can be helpful as well as injurious, as Arie van der Lugt demonstrated in his dissertation research in our group.¹⁰ Words that begin similarly to many other words are easier to extract from a preceding context than words that begin with less usual sequences; if the Dutch words *galg* ('gallows') and *geur* ('aroma') are appended to the same context (*piengalg, piengeur*), the former, which begins with the common word-initial sequence *ga-*, is spotted more readily than the latter, which begins with a rare sequence. It is as if all the activated words beginning *ga-* gang up to drag the word-initial sequence from its preceding context and thus facilitate eventual recognition of the word.

Adult listeners make very effective use of their native-language knowledge in listening. Words supported by the input are automatically activated and become available for recognition, even on the basis of only partial information; but as additional phonetic information arrives, which distinguishes between the competing candidates, it is used immediately and effectively. Competition

processes allow unwanted candidates to be quickly got rid of, and assist in the segmentation of words from context. Phoneme sequence restrictions and transition probabilities are effectively exploited. The result is the subjectively effortless process of listening in the native language.

Non-native listening

It is not unreasonable to assume that the basic processes of non-native listening are exactly those that characterize native listening: automatic multiple activation, inter-word competition. Thus, when we listen to spoken input in our second language, essentially the same thing happens as when we listen to our first: all the words which, intentionally or otherwise, receive support from the incoming speech stream become available for recognition and engage in a general process of competition. The outcome of this process is recognition of the string of words that the speaker intended to be heard.

This assumes, of course, that the words in question are stored in our vocabularies in the first place. But is this in fact true of non-native listeners? It is, after all, also reasonable to assume that our non-native vocabularies are less well stocked than our native vocabularies. For a native Dutch listener, the input *luisteraar* may precipitate competition from *lui*, *ui*, *luis*, *luister*, *te*, *aar* and *raar*, but for a listener to Dutch as a second language there may be fewer competing words. For an English native listener, recognition of *ghost writer* may involve overcoming competition from *go*, *oast*, *try*, *trite*, *rye/wry*, *right/rite* and *eye*, but, again, a non-native listener might not know all of these words and might therefore have less competition to overcome.

Since competition can, in many cases, inhibit recognition, it might seem that listening to a non-native language should, in principle, be easier than listening to the native language; at least the words one does know should be easier to recognize since they will be subjected to less interference from concurrently activated competitors. Obviously, however, this is not the case. A series of studies in our laboratory has addressed lexical processing in non-native listening in an attempt to understand why.

For example, Andrea Weber used the word-spotting task to compare recognition of English words by native listeners and skilled non-native listeners whose first language was German. Her study¹¹ focused on the role played by phoneme sequence constraints; as we saw above, these are apprehended by infant listeners, and also play a role in adult listening. English listeners, Weber found, detected words such as *lunch* more easily in contexts like *moysh-* than in contexts like *moyce-*. Because no native English words begin or end *shl-*, *moyshlunch* will activate no competitors overlapping with *lunch*, and *lunch* will be relatively easy to detect. Many English words, however, begin *sl-* (*sleep*, *slick*, *slog*, *slump*, *slow*

etc), so that *moycelunch* will presumably activate substantial numbers of potential competitors for the onset of the word, making recognition proportionately slower.

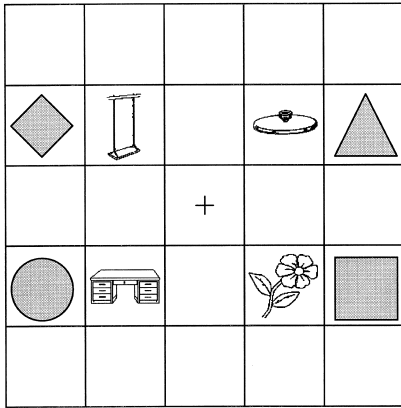
In German, interestingly, the sequence probabilities are reversed. No German words begin or end with [sl], but many begin *shl-* (*Schlaf*, *Schloss*, *schlecht*, *Schluck*, *Schlange* etc). Weber found that highly proficient German speakers of English were faster detecting words like *lunch* in *moycelunch* than in *moyshlunch*. In other words, their native German phoneme-sequence constraints were more strongly in operation with the English input than the relevant English constraints were. Thus, one factor that renders non-native listening less efficient than native listening appears to be that native phoneme sequence probabilities are inappropriately applied to non-native input.

This effect may arise because the relevant portion of the native vocabulary is actually activated by the non-native input. In a further investigation of non-native listening, Weber used the eye-tracking method to examine word activation in Dutch listeners hearing English sentences. Given the display in Figure 5(a), for instance, listeners would be instructed to *click on the desk and move it on top of the diamond*. Besides the desk, the display contains a lid, a flower and a swing. At the point at which the Dutch listeners heard *click on the de-* (see Figure 5(b)), they looked as often at the lid as at the desk (but they looked less often towards the swing or the flower). This behaviour could only mean that their native vocabulary was interfering with their non-native listening – the Dutch name for a lid is *deksel*, which begins similarly to *desk*. Even though the subjects in this experiment heard only English input, and were aware that only English would be spoken, they were apparently unable to inhibit activation of their Dutch vocabulary. In other words, one reason why non-native listening is less easy than it might be is that the native vocabulary is providing added competition.¹²

Subsequent experiments using the same approach examined the efficiency with which Dutch listeners could distinguish English phonetic contrasts to resolve potential competition. Given the display in Figure 6(a), for instance, and the instruction *click on the parrot and move it* etc, the portion *click on the pa-* should be enough to enable listeners to reject the pirate in favour of the parrot. Dutch listeners were easily able to do this, as Figure 6(b) shows; their gaze went most often to the parrot, significantly more than to any of the other three alternatives, which did not differ. Figure 7(a), however, requires them to distinguish a paddle from a pedal. In this case, the results (Figure 7(b)) showed that the discrimination was not successfully made – when the instruction referred to the paddle, the paddle and the pedal receive an equal number of looks at the point at which the distinguishing vowel information *click on the pa-* was presented (again the dissimilar alternatives receive fewer).

In the *parrot–pirate* case, the nearest Dutch vowel categories for the two

A.



B.

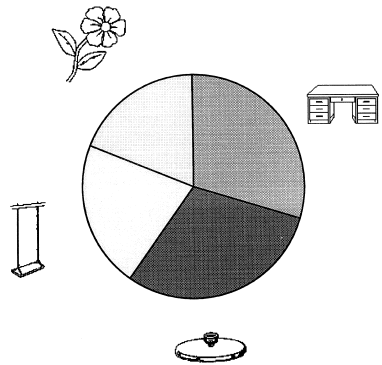
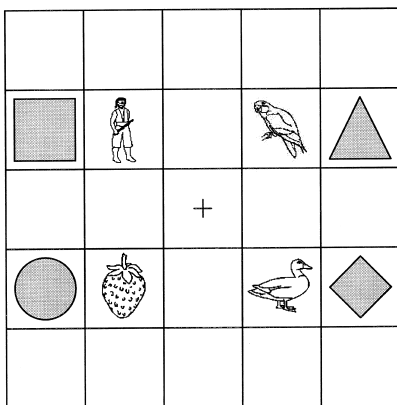


Figure 5. (a) Stimulus display for eyetracking experiment (Dutch listeners, English input); (b) Percentage of eye fixations at the point that listeners have heard *click on the de-* (results from Weber¹²).

English vowels are different. Thus, if Dutch listeners were perceiving the English vowels with reference to the contrasts relevant for distinguishing Dutch vowels, the vowels of *parrot* and *pirate* would be distinct. This is not however the case with the vowels of *paddle* and *pedal*, which assimilate to a single Dutch category. These experiments thus suggest that non-native word recognition is further

A.



B.

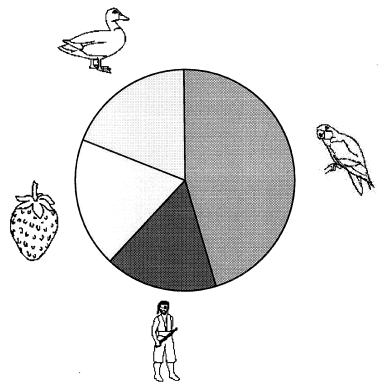
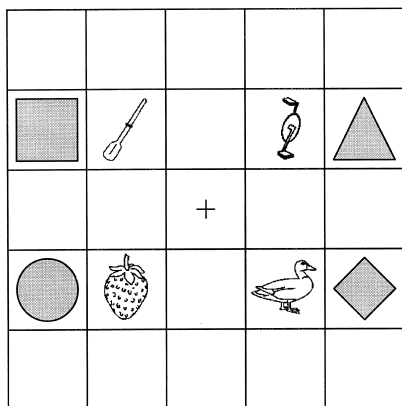


Figure 6. (a) Stimulus display for eyetracking experiment (Dutch listeners, English input); (b) Percentage of eye fixations at the point that listeners have heard *click on the pa-*.

A.



B.

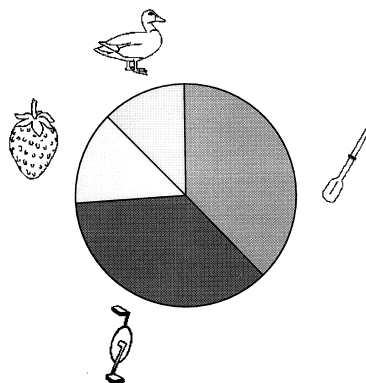


Figure 7. (a) Stimulus display for eyetracking experiment (Dutch listeners, English input); (b) Percentage of eye fixations at the point that listeners have heard *click on the pa-*.

complicated by phonetic discrimination behaviour more appropriate to the native than to the non-native vocabulary.

This is still not the whole story. The result of insufficiently precise phonetic discrimination can actually be hallucinatory word recognition. Non-native listeners can perceive words where, to native ears, no words are to be heard. Mirjam Broersma discovered this by using one of the simplest tasks with which psycholinguists examine spoken-word recognition, namely lexical decision. In a lexical decision experiment, listeners have to decide whether spoken inputs are real words or not. Hearing, for instance, *sleep* they should respond YES, hearing *skelth* they should respond NO; as in word-spotting, both their response time and their accuracy can be measured.

Figure 8 shows the response patterns of Dutch listeners in Broersma's study.¹³ All of her listeners were highly proficient in English, as can be seen from their high accuracy with real words such as *sleep* or *comfort* (to which they almost always responded YES) and clear non-words such as *crog* or *imptic* (to which they almost always responded NO). The category 'near words', however, consisted of non-words that contained phonetic discriminations required by English phonology but not by Dutch phonology. For instance, *stemp* differs from *stamp* in the same way as *pedal* from *paddle*, and as the eye-tracking study described above showed, Dutch listeners cannot easily discriminate such a pair. Likewise, *cheece*

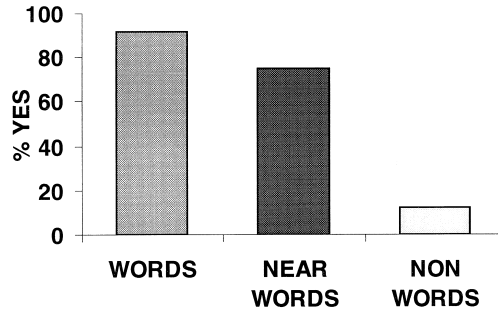


Figure 8. Percentage YES responses by Dutch listeners given English input in a lexical decision experiment, for real words, near words and non-words respectively.

differs from *cheese* in a word-final contrast (unvoiced versus voiced) that never occurs in Dutch minimal word pairs; Dutch listeners thus never have to make this kind of discrimination in word-final position. The correct response to every member of the ‘near word’ category would have been NO; but the listeners in Broersma’s experiment overwhelmingly responded YES. For these listeners, *stemp* and *cheece* are effectively English words.

This may not seem to be a severe problem for non-native listening – after all, native speakers do not usually deliberately mispronounce words. In general, the words that non-native listeners hear will be real words and not near words. But recall that word candidates are automatically activated when phonetic interpretation of the input supports them, even when only partial support is present or when the words are only spuriously present in the input via embedding within or across the actually spoken words. Such near words may often be present in spoken input. For instance, the English word *stem* will receive competition from the word *stamp*, for any listener who does not distinguish the two vowels. Likewise, the movie title *The Last Emperor* will, for such a listener, activate the word *stamp* as a competing embedded string. The phrase *each Easter* similarly contains an embedded string that will be heard as *cheese* by a listener who fails to make the word-final voicing discrimination.

The problem is, of course, not confined to non-native listeners to English. Non-native listeners to Dutch notoriously have difficulty with the vowel contrast in *uit* versus *oud*, so that the word *oud* might be activated as a spurious embedding when non-native listeners hear *geluid*. And any pair of languages in which the required sets of phonetic discriminations are not exactly identical would produce similar phenomena.

Activation and competition in non-native listening are thus not simply determined by the size of the listener’s vocabulary in the second language. The

sensitivity to phonetic contrasts also plays a role, and can result in spurious competitors, which should not be activated for native listeners. And the native vocabulary also provides additional unnecessary competition. Non-native listening can thus involve far more activation (and consequent competition) than native listening.

Conclusions

Infant listening is productively employed. Knowledge of the phoneme contrasts used in the environmental language, of the relative likelihood of phonetic sequences, and of useful techniques to extract words from running speech contexts, is all built up by infants before effective spoken communication begins. These types of knowledge contribute to the array of language-specific procedures that constitute native listening: the attuning of the linguistic processing system to a particular language in the interests of processing efficiency. Adult listeners, although they can exploit the resources of a well-stocked vocabulary in comprehending speech, also draw upon the very dimensions of language structure, which were constructed by the infant. Non-native listening, however, is in turn hindered by the efficiency with which the native procedures operate; even when native procedures are ill-suited to the structure of a non-native language, their application is difficult to inhibit. The result can be an increase in spurious word activation and competition, making word recognition in the second language more difficult. The baby's hard work in the first year of life pays off handsomely in native listening efficiency; drawbacks only become apparent when adults also engage in non-native listening.

Acknowledgements

This material was presented as a Heisenberg lecture to the Carl Friedrich von Siemens Stiftung, München, in March 2001, as the Diesrede of the Katholieke Universiteit Nijmegen in May 2001, and to the 13th Annual Scientific Meeting of the Academia Europaea in Rotterdam in June 2001. The Dutch version of the May presentation was published as *'De baby in je hoofd: Luisteren naar eigen en andermans taal'* by Nijmegen University Press. Many thanks to the colleagues with whom the research reported here was undertaken, in particular James McQueen, and the five PhD students whose work is described: Joost van de Weijer, Arie van der Lugt, Andrea Weber, Anne-Pier Salverda and Mirjam Broersma. International collaborations with the groups of the late Peter Jusczyk (Johns Hopkins University) and Janet Werker (University of British Columbia) are also gratefully acknowledged.

References

1. P. W. Jusczyk, P. A. Luce and J. Charles-Luce (1994) Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, **33**, 630–645.
2. P. W. Jusczyk, A. Cutler and N. Redanz (1993) Infants' preference for the predominant stress patterns of English words. *Child Development*, **64**, 675–687.
3. J. van de Weijer (1999) Language input for word discovery. PhD thesis, University of Nijmegen.
4. P. W. Jusczyk and R. N. Aslin (1995) Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology*, **29**, 1–23.
5. D. M. Houston, P. W. Jusczyk, C. Kuijpers, R. Coolen and A. Cutler (2000) Cross-language word segmentation by 9-month-olds. *Psychonomic Bulletin and Review*, **7**, 504–509.
6. I. Maddieson (1984) *Patterns of Sounds* (Cambridge: Cambridge University Press).
7. A.-P. Salverda (in preparation) PhD thesis, University of Nijmegen.
8. D. G. Norris, J. M. McQueen and A. Cutler (1995) Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **21**, 1209–1228.
9. J. McQueen (1998) Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, **39**, 21–46.
10. A. van der Lugt (1999) From speech to words. PhD thesis, University of Nijmegen.
11. A. Weber (submitted) Phonotactic constraints and the segmentation of native and non-native spoken language.
12. A. Weber (2001) Language-specific listening: The case of phonetic sequences. PhD thesis, University of Nijmegen.
13. M. Broersma (in preparation) PhD thesis, University of Nijmegen.
14. J. F. Werker and C. E. Lalonde (1988) Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, **24**, 672–683.

About the Author

Anne Cutler is one of four directors of the Max-Planck-Institute for Psycholinguistics in Nijmegen, the Netherlands, where she is responsible for work in the area of language comprehension. Previously she was at the University of Sussex and the MRC Applied Psychology Unit in Cambridge. Her work centres on the recognition of spoken language, in particular the role of prosody in comprehension, and this in turn led her into cross-linguistic work, with current projects involving over a dozen languages from different families.